

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 45.463 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771  
Sequence: 1 MKKKLHHHHHTSAGITR.....TTWTSAMRHPOGKXXX 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1771	100.0	334	5	ABG32182 HCV prote
2	1660	93.7	409	5	ABG32181 HCV prote
3	1589	89.7	303	5	ABG32183 HCV prote
4	1589	89.7	341	5	ABG32187 HCV prote
5	1589	89.7	352	5	ABG32186 HCV prote
6	1589	89.7	380	5	ABG32185 HCV prote
7	1589	89.7	393	5	ABG32184 HCV prote
8	1580	89.2	303	5	ABG32191 HCV prote
9	1579	89.2	303	5	ABG32189 HCV prote
10	1570	88.7	301	5	ABG32190 HCV prote
11	1532	86.5	292	5	ABG32188 HCV prote
12	1531	86.4	2201	5	ABG30591 Hepatitis
13	1531	86.4	2201	5	ABG30592 Hepatitis
14	1531	86.4	2201	5	ABG30593 Hepatitis
15	1531	86.4	2201	5	ABG30581 Hepatitis
16	1531	86.4	2201	5	ABG30593 Hepatitis
17	1531	86.4	2201	5	ABG30582 Hepatitis
18	1531	86.4	2201	5	ABG30580 Hepatitis
19	1531	86.4	2201	5	ABG30587 Hepatitis
20	1531	86.4	2201	5	ABG30599 Hepatitis
21	1531	86.4	2201	5	ABG30594 Hepatitis
22	1531	86.4	2201	5	ABG30598 Hepatitis
23	1531	86.4	2201	5	ABG30595 Hepatitis
24	1531	86.4	3010	5	ABG32458 Hepatitis
25	1531	86.4	3010	5	ABG32459 Hepatitis

26	1531	86.4	3010	5	ABG32451 Hepatitis
27	1531	86.4	3010	5	ABG32455 Hepatitis
28	1531	86.4	3010	5	ABG32457 Hepatitis
29	1531	86.4	3010	5	ABG32460 Hepatitis
30	1531	86.4	3010	5	ABG32461 Hepatitis
31	1531	86.4	3010	5	ABG32454 Hepatitis
32	1531	86.4	3010	5	ABG32456 Hepatitis
33	1531	86.4	3011	5	ABG32456 Hepatitis
34	1528	86.3	2201	5	ABG30586 Hepatitis
35	1528	86.3	2201	5	ABG30589 Hepatitis
36	1528	86.3	2201	5	ABG30583 Hepatitis
37	1528	86.3	2201	5	ABG30588 Hepatitis
38	1528	86.3	2307	3	AA770064 Recombina
39	1528	86.3	3010	2	AA770064 Recombina
40	1528	86.3	3010	2	AA770064 Recombina
41	1527	86.2	2201	5	ABG30590 Hepatitis
42	1525	86.1	2307	3	AA770065 Recombina
43	1525	86.1	3010	5	ABG32452 Hepatitis
44	1524	86.1	2201	5	ABG30584 Hepatitis
45	1524	86.1	2201	5	ABG30602 Hepatitis

## ALIGNMENTS

RESULT 1	ABG32182	standard; protein; 334 AA.
ID	ABG32182	
XX	ABG32182;	
AC	ABG32182;	
XX	05-NOV-2002 (first entry)	
DT	05-NOV-2002	
XX	HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.	
DE	HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.	
XX	HCV, enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;	
KW	chronic liver disease; cirrhosis; end-stage liver disease; virulence;	
KW	hepatotropic; antiinflammatory; leucylalanylalanine oxide; LDMO;	
KW	chaotropic agent; 4K-6H (904-1206)st-4K; mutant; mutain.	
XX	Hepatitis C virus.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..15
FT	Protein	/note="4-Lys/His tag"
FT		16..302
FT		/note="Truncated NS2/3 protease"
FT	Peptide	319..334
FT		/note="Streptavidin/4-Lys tag"
XX	WO200248375-A2.	
XX	20-JUN-2002.	
PD	13-DEC-2001; 2001WO-CA001796.	
PF	15-DEC-2000; 2000US-0256031P.	
XX		
PR	(BOEH) BOEHRINGER INGELHEIM CANADA LTD.	
PA		
XX	Thibault D, Lamarte R, Pilote L, Pause A;	
PI	WPI; 2002-599511/64.	
XX	N-PSDB; ABR90407.	
DR		
XX	Novel polypeptide for screening inhibitors of non-structural proteases	
PT	useful as therapeutic agents against hepatitis C virus, comprises full	
PT	length non-structural protease, or its truncation.	
XX	Claim 39; Fig 9B; 67bp; English.	
PS		
XX	The invention relates to an isolated polypeptide consisting of a full-	

length of HCV hepatitis C virus) non-structural (NS2/3 protease (referred to also as NS2/3(810-1266)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residue 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldimethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as AB832198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and producing cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of, or active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 4K-5H (904-1206)-E1-K comprising a truncated NS2/3 protein with a four Lys/His N-terminal tag, a C-terminal streptavidin tag and C-terminal four Lys tag

DE	HCV protease NS2/3 (810-1206) .	
XX		
XX	HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;	
KW	chronic liver disease; cirrhosis; end-stage liver disease; viraemia;	
KW	hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;	
KW	chaotropic agent; mutant; mutagen.	
XX		
OS	Hepatitis C virus.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	388..409
FT		/note "Streptavidin tag"
XX		
FN	WO200248375-A2.	
XX		
PD	20-JUN-2002.	
XX		
PF	13-DEC-2001; 2001WO-CA001796.	
XX		
PR	15-DEC-2000; 2000US-0256031P.	
XX		
PA	(BOEH) BOEHRINGER INGELHEIM CANADA LTD.	
XX		
PI	Thibeault D, Lamarre D, Maurice R, Pilote L, Paus A;	
XX		
DR	WPI: 2002-599511/64.	
XX	N-PSDB; ABX90406.	
XX		
XX	Novel polypeptide for screening inhibitors of non-structural proteases	
PT	useful as therapeutic agents against hepatitis C virus, comprises full	
PT	length non-structural protease, or its truncation.	
XX		
PS	Claim 42; Fig 1B; 67pp; English.	
XX		

The invention relates to an isolated polypeptide consisting of a full-

Query Match	100.0%	Score 1771;	DB 5;	Length 334;
Best Local Similarity	100.0%	Pred. No. 5,4e-164;		
Matches 334;	Conservative 0;	Mismatches 0;	Indels 0	Gaps 0
QY	1	MKKKKLEHHHHHTTSAGITKVPYFPVPAQGLIPACMLYKKAAGHYVQMAFMKLAALTGY	60	
DB	1	MKKKKLEHHHHHTTSAGITKVPFPVPAQGLIPACMLYKKAAGHYVQMAFMKLAALTGY	60	
QY	61	VYDHLTELPQMAHAGRLDAVAVEPIPSDMEVKIITWGADTAAQGDIIISGLPVSARRG	120	
DB	61	VYDHLTELPQMAHAGRLDAVAVEPIPSDMEVKIITWGADTAAQGDIIISGLPVSARRG	120	
QY	121	ETLLGPRDNEGGGMLLAPITAYSQQRGLGCIITSLTRDKNQVEGVQVASTATOS	180	
DB	121	ETLLGPRDNEGGGMLLAPITAYSQQRGLGCIITSLTRDKNQVEGVQVASTATOS	180	
QY	181	FLATCVNGVCMVTFHGAGSGKTLAGPKPIITQWYTNVDDIVGWQAPGAASMTPTCGSS	240	
DB	181	FLATCVNGVCMVTFHGAGSKTLAGPKPIITQWYTNVDDIVGWQAPGAASMTPTCGSS	240	
QY	241	DLVLYNRHADVTEVRRRGDSRGSLLSPRVSYLTKSSGGPILTCGSAVAGFPRAAVCTRQ	300	
DB	241	DLVLYNRHADVTEVRRRGDSRGSLLSPRVSYLTKSSGGPILTCGSAVAGFPRAAVCTRQ	300	
QY	301	VAKAVDFIPVESMETTMTTSAMRHPQFGKXXX	334	
DB	301	VAKAVDFIPVESMETTMTTSAMRHPQFGKXXX	334	

RESULT 2  
 ABG32181  
 ID ABG32181 standard; protein; 409 AA.  
 AC ABG32181;

DT	05-NOV-2002	(first entry)
XX		

Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTPLOQMAHAG 75  
 DB 95 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTPLOQMAHAG 154  
 QY 76 LRDLAVAVEPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135  
 DB 155 LRDLAVAVEPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 214  
 QY 136 RLAPITRAYSQOTRGLIGCIITSLTGRDKNQVGEVQVSTATQSPLATCVNGVCTVTH 195  
 DB 215 RLAPITRAYSQOTRGLIGCIITSLTGRDKNQVGEVQVSTATQSPLATCVNGVCTVTH 274  
 QY 196 GAGSKTLAGPKPIPTQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255  
 DB 275 GAGSKTLAGPKPIPTQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 334  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 335 RRGDSRGSLLSPRPVSYLKSSGGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 394  
 QY 316 TMRSSAMRHAPQFGG 330  
 DB 395 TMRSSAMRHAPQFGG 409

## RESULT 3

ABG32183  
 ID ABG32183 standard; protein; 303 AA.

XX ABG32183;

DT 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 904-1206.

XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.

OS Hepatitis C virus.  
 XX Synthetic.

PN W0200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WC-CA001796.

FR 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-559511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.

PS Claim 39; Page 58-59; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution

CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32183; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving dialyzing refolded inactive NS2/3 protease in a medium  
 CC containing an activating detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 904-1206  
 CC (numbered relative to the full length NS2/3 protein)

SQ Sequence 303 AA;

Query Match 89.7%; Score 1589; DB 5; Length 303;

Best Local Similarity 100.0%; Pred. No. 2.7e-146;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTPLOQMAHAG 75  
 DB 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTPLOQMAHAG 60  
 QY 76 LRDLAVAVEPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135  
 DB 61 LRDLAVAVEPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 120  
 QY 136 RLAPITRAYSQOTRGLIGCIITSLTGRDKNQVGEVQVSTATQSPLATCVNGVCTVTH 195  
 DB 121 RLAPITRAYSQOTRGLIGCIITSLTGRDKNQVGEVQVSTATQSPLATCVNGVCTVTH 180  
 QY 196 GAGSKTLAGPKPIPTQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255  
 DB 181 GAGSKTLAGPKPIPTQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 240  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 241 RRGDSRGSLLSPRPVSYLKSSGGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 QY 316 TMR 318  
 DB 301 TMR 303

## RESULT 4

ABG32187  
 ID ABG32187 standard; protein; 341 AA.

XX ABG32187;

DT 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 866-1206.

XX HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.  
XX WO200248375-A2.  
XX  
XX 20-JUN-2002.  
XX  
XX 13-DEC-2001; 2001WO-CA001796.  
XX  
XX 15-DEC-2000; 2000US-0256031P.  
XX  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX  
XX WPI; 2002-599511/64.  
XX  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
XX  
XX Claim 41; Page 62-63; 67pp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
XX its truncation or a mutated sequence, where the protease is in a solution  
XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
XX appearing as ABG32186; (3) producing (M1) a refolded, inactive HCV NS2/3  
XX chaotropic agent, refolding the isolated protease in the presence of a  
XX reducing agent, and LDAO in the presence of reduced concentration of the  
XX chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
XX protease, involving diluting refolded inactive NS2/3 protease in a medium  
XX containing an activation detergent to induce auto-cleavage of the NS2/3  
XX protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
XX protease, involving incubating the active NS2/3 protease produced by M2  
XX for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
XX cleavage products or their fragments, and measuring the presence of  
XX absence of uncleaved NS2/3 protease, cleavage products or their fragments  
XX ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
XX active NS2/3 protease, involving carrying out M3 in the presence of, or  
XX absence of the potential inhibitor, comparing the amount of uncleaved  
XX NS2/3 protease, cleavage products or their fragments. The protease is  
XX useful for detailed biochemical characterisation of the enzymes and in  
XX the development of in vitro assays for screening novel inhibitors of  
XX NS2/3 protease which are useful as therapeutic agents against HCV  
XX infection (which causes chronic liver disease, cirrhosis and end-stage  
XX liver disease. M1 is useful for high level production of protease. The  
XX protein sequence represents the NS2/3 truncation mutant 866-1206  
XX (numbered relative to the full length NS2/3 protein)  
XX  
XX Sequence 341 AA;  
XX  
XX Query Match 89.7%; Score 1589; DB 5; Length 341;  
XX Best Local Similarity 100.0%; Pred. No. 3.2e-146;  
XX Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 16 AGTTKVVYFPAAGLIRACVLRKAAAGHYVQVAFMKLALTTGYVDHLPQDWAHAG 75  
XX  
XX 39 AGTTKVVYFPAAGLIRACVLRKAAAGHYVQVAFMKLALTTGYVDHLPQDWAHAG 98  
XX  
XX 76 LRDLAVAVEPFIPEFDMVEKLTITGADTAACGDIISGLPVSARGREILIGPADNFGQGM 135  
XX  
XX 99 LRDLAVAVEPFIPEFDMVEKLTITGADTAACGDIISGLPVSARGREILIGPADNFGQGM 158  
XX  
XX 136 RLAPITAVSQGTGRLIGCTITSLTGRDKNVGEGVAVSTAQSLFATCVNGVCTVFR 195  
XX  
XX 159 RLAPITAVSQGTGRLIGCTITSLTGRDKNVGEGVAVSTAQSLFATCVNGVCTVFR 218

QY 196 GAGSKTLAAGPKPTTQWNTVNDOLVGMQAPRGARSTPCTCGSSDLVYTRRADVTPVR 255  
DB 219 GAGSKTLAAGPKPTTQWNTVNDOLVGMQAPRGARSTPCTCGSSDLVYTRRADVTPVR 278  
QY 256 RRGDSRGSLSFPRFVSYLKSGSGGLTLCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 315  
DB 279 RRGDSRGSLSFPRFVSYLKSGSGGLTLCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 338  
QY 316 TMR 318  
DB 339 TMR 341  
RESULT 5  
ABG32186  
ID ABG32186 standard; protein, 352 AA.  
XX  
XX ABG32186;  
XX  
XX 05-NOV-2002 (first entry)  
XX  
XX HCV protease NS2/3 truncation mutant 855-1206.  
XX  
XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
XX chaotropic agent; mutant; mutein.  
XX  
XX Hepatitis C virus.  
XX Synthetic.  
XX  
XX WO200248375-A2.  
XX  
XX 20-JUN-2002.  
XX  
XX 13-DEC-2001; 2001WO-CA001796.  
XX  
XX 15-DEC-2000; 2000US-0256031P.  
XX  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX  
XX WPI; 2002-599511/64.  
XX  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
XX  
XX Claim 41; Page 61-62; 67pp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
XX its truncation or a mutated sequence, where the protease is in a solution  
XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
XX appearing as ABG32186; (3) producing (M1) a refolded, inactive HCV NS2/3  
XX chaotropic agent, refolding the isolated protease in the presence of a  
XX reducing agent, and LDAO in the presence of reduced concentration of the  
XX chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
XX protease, involving diluting refolded inactive NS2/3 protease in a medium  
XX containing an activation detergent to induce auto-cleavage of the NS2/3  
XX protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
XX protease, involving incubating the active NS2/3 protease produced by M2  
XX for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
XX cleavage products or their fragments, and measuring the presence of



CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 855-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 CC  
 CC Sequence 352 AA:

Query Match 89.7%; Score 1589; DB 5; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFKMLAALTGYVDHLTPLODMAHG 75  
 DB 50 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFKMLAALTGYVDHLTPLODMAHG 109  
 QY 76 LRDLAVALVEPIFSDMEVKITMGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 135  
 DB 110 LRDLAVALVEPIFSDMEVKITMGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 169  
 QY 136 RLALPITAVSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTAFATCNGVCMVTFH 195  
 DB 170 RLALPITAVSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTAFATCNGVCMVTFH 229  
 QY 196 GAGSKTLAGKSPITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLTRADVIPIVR 255  
 DB 230 GAGSKTLAGKSPITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLTRADVIPIVR 289  
 QY 256 RRGDSRGSLLSPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 315  
 DB 290 RRGDSRGSLLSPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 349  
 QY 316 TMR 318  
 DB 350 TMR 352

RESULT 6

ABG32185  
 ID ABG32185 standard; protein; 380 AA.

AC ABG32185;  
 DT 05-NOV-2002 (first entry)  
 XX

DE HCV protease NS2/3 truncation mutant 827-1206.

KW HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
 KW chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KW hepatocellular carcinoma; leucyldiethylamine oxide; LDAO;  
 KW chaotropic agent; mutant; mutagen.

OS Hepatitis C virus.  
 OS Synthetic.

PN WO200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WO-CA001796.

PR 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamare R, Maurice R, Pilote L, Paus A;  
 PI

XX WPI, 2002-599511/64.  
 DR Novel polypeptide for screening inhibitors of non-structural proteases  
 XX useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 41; Page 60-61; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS2/3) protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32185; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 827-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 CC  
 CC Sequence 380 AA:

Query Match 89.7%; Score 1589; DB 5; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFKMLAALTGYVDHLTPLODMAHG 75  
 DB 78 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFKMLAALTGYVDHLTPLODMAHG 137  
 QY 76 LRDLAVALVEPIFSDMEVKITMGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 135  
 DB 138 LRDLAVALVEPIFSDMEVKITMGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 197  
 QY 136 RLALPITAVSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTAFATCNGVCMVTFH 195  
 DB 198 RLALPITAVSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTAFATCNGVCMVTFH 257  
 QY 196 GAGSKTLAGKSPITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLTRADVIPIVR 255  
 DB 258 GAGSKTLAGKSPITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLTRADVIPIVR 317  
 QY 256 RRGDSRGSLLSPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 315  
 DB 318 RRGDSRGSLLSPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 377  
 QY 316 TMR 318  
 DB 378 TMR 380

RESULT 7  
 ABG32184  
 ID ABG32184 standard; protein; 393 AA.  
 AC  
 XX ABG32184;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE HCV protease NS2/3 truncation mutant 815-1206.  
 XX  
 XX HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocid;  
 XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX WO200248375-A2.  
 XX  
 PD 20-JUN-2002.  
 XX  
 PF 13-DEC-2001; 2001WO-CA001796.  
 XX  
 PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Thibeault D, Lamerre D, Maurice R, Pilote L, Pause A;  
 XX  
 DR WPI; 2002-599511/64.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 PS Claim 41; Page 59-60; 67pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32184; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC chaotropic agent, isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 XX

SO Sequence 393 AA:  
 Query Match 89.7%; Score 1589; DB 5; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 4e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 AGITKVPYFVRAQGLIRACMLVRKAAAGHYVOMAFMKLAALGTGYVYDHLTPLOMAHAG 75  
 DB 91 AGITKVPYFVRAQGLIRACMLVRKAAAGHYVOMAFMKLAALGTGYVYDHLTPLOMAHAG 150  
 QY 76 LRDLAAVAVPVIKSDMEVXIIITWGDPTAACGDIISGLPVSAARGREIILGPADNFEQGM 135  
 DB 151 LRDLAAVAVPVIKSDMEVXIIITWGDPTAACGDIISGLPVSAARGREIILGPADNFEQGM 210  
 QY 136 RLAPITTAASQOTRGLGCIITSLGRDKNQVGEVQVYSTATQSFATCVCNVGCMYTFH 195  
 DB 211 RLAPITTAASQOTRGLGCIITSLGRDKNQVGEVQVYSTATQSFATCVCNVGCMYTFH 270  
 QY 196 GAGSKTLAAGPKPITOMYTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPVR 255  
 DB 271 GAGSKTLAAGPKPITOMYTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPVR 330  
 QY 256 RRGDSRGSLLSRPPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAAVDFIPVESMET 315  
 DB 331 RRGDSRGSLLSRPPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAAVDFIPVESMET 390  
 QY 316 TMR 318  
 DB 391 TMR 393  
 RESULT 8  
 ABG32191  
 ID ABG32191 standard; protein; 303 AA.  
 AC  
 XX ABG32191;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE HCV protease NS2/3 truncation 904-1206/Cys993A1A.  
 XX  
 CC HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 CC chronic liver disease; cirrhosis; end-stage liver disease; virocid;  
 CC hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 CC chaotropic agent; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 90 /note= "wild-type Cys substituted by Ala"  
 FT  
 XX MO200248375-A2.  
 XX  
 PD 20-JUN-2002.  
 XX  
 PF 13-DEC-2001; 2001WO-CA001796.  
 XX  
 PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Thibeault D, Lamerre D, Maurice R, Pilote L, Pause A;  
 XX  
 DR WPI; 2002-599511/64.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 PS Disclosure; Page 65-66; 67pp; English.  
 XX



Best Local Similarity 99.7%; Pred. No. 2.6e-145; Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75
DB 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 60
QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRRREITLGPADNFEQGM 135
DB 61 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRRREITLGPADNFEQGM 120
QY 136 RLAPITAVSQOTRGLGCIITSLTGRDNQVEGVQVSTATOSFLATCVNGVMTVFH 195
DB 121 RLAPITAVSQOTRGLGCIITSLTGRDNQVEGVQVSTATOSFLATCVNGVMTVFH 180
QY 196 GAGSKTLGPKPITOMTNTDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVPVR 255
DB 181 GAGSKTLGPKPITOMTNTDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVPVR 240
QY 256 RRGDSRGSILSPRPVSYLKSGSGGFLCPSGHAAGIFPAAVCTRGVAKAVDFIVESMET 315
DB 241 RRGDSRGSILSPRPVSYLKSGSGGFLCPSGHAAGIFPAAVCTRGVAKAVDFIVESMET 300
QY 316 TMR 318
DB 301 TMR 303

```

## RESULT 10

ABG32190  
ID ABG32190 standard; protein; 301 AA.

AC ABG32190;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation 904-1206/deltaLeu1026-Ala1027.

KW HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;  
 KW hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KW chaotropic agent; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

FX Key Location/Qualifiers

FT Misc-difference 122.123

FT /note= "Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"

FN W0200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WO-CA001796.

PR 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

DR WPI; 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.

PS Example 7; Page 64-65; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal

CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation 904-1206 mutant  
 CC deltaLeu1026-Ala1027 (numbered relative to the full length NS2/3 protein)  
 CC a mutant devoid of autocatalytic activity

CC Sequence 301 AA;

CC Query Match 88.7%; Score 1570; DB 5; Length 301;

CC Best Local Similarity 99.3%; Pred. No. 1.9e-144; Matches 301; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75
DB 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 60
QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRRREITLGPADNFEQGM 135
DB 61 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRRREITLGPADNFEQGM 120
QY 136 RLAPITAVSQOTRGLGCIITSLTGRDNQVEGVQVSTATOSFLATCVNGVMTVFH 195
DB 121 RLAPITAVSQOTRGLGCIITSLTGRDNQVEGVQVSTATOSFLATCVNGVMTVFH 178
QY 196 GAGSKTLGPKPITOMTNTDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVPVR 255
DB 179 GAGSKTLGPKPITOMTNTDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVPVR 238
QY 256 RRGDSRGSILSPRPVSYLKSGSGGFLCPSGHAAGIFPAAVCTRGVAKAVDFIVESMET 315
DB 239 RRGDSRGSILSPRPVSYLKSGSGGFLCPSGHAAGIFPAAVCTRGVAKAVDFIVESMET 298
QY 316 TMR 318
DB 299 TMR 301

```

## RESULT 11

ABG32188  
ID ABG32188 standard; protein; 292 AA.

AC ABG32188;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 915-1206.

DB

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by concentrating it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 915-1206  
 CC (numbered relative to the full length NS2/3 protein)

Query Match	100.0%	Score 1532	DB 5	Length 292
Best local similarity	100.0%	Pred. No. 9,5e-141		
Matches	292	Conservative	0	Mismatches 0
			Indels	0
			Gaps	0
QY	27	AGGIRACM.LVRKAGGHYQMAFMKLAALGTGYVDHILTPDPAHAAGLADLAIVANPEV	86	
DB	1	AGGIRACM.LVRKAGGHYQMAFMKLAALGTGYVDHILTPDPAHAAGLADLAIVANPEV	60	
QY	87	IFSDMEKIIITGWGADTLACGDIISGHPVARRRREITLLGPADNFEFGQCMRLAPITAYSQ	146	

Db 61 IFSDEWKLITWGADPLAACGDIISGLPVASARGREILLGPADNPEGQWRLAPITAYSQ 130

Qy 147 QTFGLIGCIITSITGPDKNQVGEVQVNSTANQSLACNNGVQMTYFHAGSGKTLAQP 206

Db 121 QTFGLIGCIITSITGPDKNQVGEVQVNSTANQSLACNNGVQMTYFHAGSGKTLAQP 180

Qy 207 GPITQMTYTVNDODLVGWQAPPGARSKPTCTGSSDLVYVTRHADVYIVRRRRDSGSL 266

Db 181 GPITQMTYTVNDODLVGWQAPPGARSKPTCTGSSDLVYVTRHADVYIVRRRRDSGSL 240

Qy 267 PRIVSTLTKSSGGPILCPESGNAVGIIFRAVACPRGAKAVNDIPIVESMTYLR 318

Db 241 PRIVSTLTKSSGGPILCPESGNAVGIIFRAVACPRGAKAVNDIPIVESMTYLR 292

ID	ABG30601	standard; protein; 2201 AA.
XX	ABG30601,	
XX	21-OCT-2002	(first entry)
XX	Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.	
XX	Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor;	
XX	cell culture replication; NS2/3; NS3/4; NS3, NS5B; mutant; mutein.	
XX	Hepatitis C virus.	
XX	Synthetic.	
XX	Key	Location/Qualifiers
XX	Misc-difference 882	/label= Arg, Lys
XX	Misc-difference 2183	/note= "Wild type Met substituted by Thr"
XX	MO200252015-A2.	
XX	04-JUL-2002.	
XX	20-DEC-2001; 2001MO-CA001843.	
XX	22-DEC-2000; 2000US-0257857P.	
XX	(BOEH ) BOEHRINGER INGELHEIM CANADA LTD.	
XX	Kukolj G, Pause A;	
XX	WPI; 2002-575382/61.	
XX	New self-replicating RNA molecules from Hepatitis C virus (HCV), which	
XX	possess enhanced transduction or replication efficiency, useful for	
XX	evaluating potential inhibitors of HCV replication.	
XX	Claim 3; Page; 140pp; English.	
XX	The invention describes a self-replicating hepatitis C virus (HCV)	
XX	polynucleotide molecule comprising a 5'-non translated region (NTR),	
XX	where guanine at position 1 is substituted for adenine, a HCV polypeptide	
XX	region coding for a HCV polypeptide; and a 3'-NTR region. The self-	
XX	replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating	
XX	potential inhibitors of HCV replication. The HCV RNA molecule is also	
XX	useful for efficiently establishing cell culture replication. The self-	
XX	replicating polynucleotide molecule contains a 5'-NTR, where G at	
XX	position 1 is substituted for A, and therefore provides an alternative to	
XX	existing systems comprising a self-replicating HCV RNA molecule that, in	
XX	conjunction with mutations in the HCV non-structural region, such as the	
XX	G(7042)C/R mutations, transduces and/or replicates with greater	
XX	efficiency. This amino acid sequence represents a mutant of the hepatitis	
XX	C virus replicon AGK12 and contains the viral protease NS2/3, protease	
XX	complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:	

CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

SQ Sequence 2201 AA:

Query Match 86.4%; Score 1531; DB 5; Length 2201;  
 Best Local Similarity 94.7%; Pred. No. 2.1e-139;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```

QY 16 AGITKVPFVFAAGLIRACMLVRKAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 75
DB 95 AGITKVPFVFAAGLIRACMLVRKAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 154
QY 76 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135
DB 155 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 214
QY 136 RLAPITAYVSOQTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 195
DB 215 RLAPITAYVSOQTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 274
QY 196 GAGSKTLAGPKGPIITQMTNTVDQVLVGWQAPPGARSMTPTCGSSDLYLTVRHADVTPVR 255
DB 275 GAGSKTLAGPKGPIITQMTNTVDQVLVGWQAPPGARSMTPTCGSSDLYLTVRHADVTPVR 334
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIVESHMET 315
DB 335 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIVESHMET 394
QY 316 TWRT 319
DB 395 TWRS 398

```

#### RESULT 13

ABG30591  
 ID ABG30591 standard; protein; 2201 AA.

AC ABG30591;

DT 21-OCT-2002 (first entry)

DB Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 751 /note= "Wild type Ser substituted by Gly"

FT Misc-difference 882 /label= Arg, Lys

XX

XX NO200252015-A2.

XX 04-JUL-2002.

XX 20-DEC-2001; 2001WO-CA001843.

XX 22-DEC-2000; 2000US-0257857P.

XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.

XX Kukulj G, Pause A;

XX

XX WPI; 2002-575382/61.

XX

XX

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for

PT evaluating potential inhibitors of HCV replication.

XX Claim 3; Page; 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G12042/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon APCK12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

SQ Sequence 2201 AA:

Query Match 86.4%; Score 1531; DB 5; Length 2201;  
 Best Local Similarity 94.7%; Pred. No. 2.1e-139;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```

QY 16 AGITKVPFVFAAGLIRACMLVRKAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 75
DB 95 AGITKVPFVFAAGLIRACMLVRKAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 154
QY 76 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135
DB 155 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 214
QY 136 RLAPITAYVSOQTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 195
DB 215 RLAPITAYVSOQTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 274
QY 196 GAGSKTLAGPKGPIITQMTNTVDQVLVGWQAPPGARSMTPTCGSSDLYLTVRHADVTPVR 255
DB 275 GAGSKTLAGPKGPIITQMTNTVDQVLVGWQAPPGARSMTPTCGSSDLYLTVRHADVTPVR 334
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIVESHMET 315
DB 335 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIVESHMET 394
QY 316 TWRT 319
DB 395 TWRS 398

```

#### RESULT 14

ABG30600  
 ID ABG30600 standard; protein; 2201 AA.

AC ABG30600;

DT 21-OCT-2002 (first entry)

DB Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 882 /label= Arg, Lys

XX



FT Misc-difference 1357  
/note= "Wild type Pro substituted by Leu"  
XX  
XX WO200252015-A2.  
XX  
XX 04-JUL-2002.  
XX  
XX 20-DEC-2001; 2001WO-CA001843.  
XX  
XX 22-DEC-2000; 2000US-0257857P.  
XX  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
XX Kukulj G, Pause A;  
XX  
XX WPI; 2002-575382/61.  
XX  
XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.  
XX  
XX Claim 3; Page: 140pp; English.  
XX  
XX The invention describes a self-replicating hepatitis C virus (HCV)  
CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
CC useful for efficiently establishing cell culture replication. The self-  
CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
CC position 1 is substituted for A, and therefore provides an alternative to  
CC existing systems comprising a self-replicating HCV RNA molecule that, in  
CC conjunction with mutations in the HCV non-structural region, such as the  
CC G12042/C/R mutations, transduces and/or replicates with greater  
CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
CC C virus replicon Apk12 and contains the viral protease NS2/3, protease  
CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
CC This sequence does not appear in the specification but has been created  
CC from the wild type sequence shown in ABG30580 using information given in  
CC the claims of the invention  
XX  
XX Sequence 2201 AA;  
SQ  
Query Match 86.4%; Score 1531; DB 5; Length 2201;  
Best Local Similarity 94.7%; Pred. No. 2,1e-139;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

RESULT 15  
ID ABG30581 standard; protein; 2201 AA.  
XX  
XX ABG30581;  
AC  
XX  
XX 21-OCT-2002 (first entry)  
DT  
XX  
XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.  
DE  
XX  
XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
XX cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
XX  
XX Hepatitis C virus.  
XX  
XX WO200252015-A2.  
XX  
XX 04-JUL-2002.  
XX  
XX 20-DEC-2001; 2001WO-CA001843.  
XX  
XX 22-DEC-2000; 2000US-0257857P.  
XX  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
XX Kukulj G, Pause A;  
XX  
XX WPI; 2002-575382/61.  
XX  
XX N-PSDB; ABK6573.  
XX  
XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.  
XX  
XX Disclosure; Page 49-58; 140pp; English.  
XX  
XX The invention describes a self-replicating hepatitis C virus (HCV)  
CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
CC useful for efficiently establishing cell culture replication. The self-  
CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
CC position 1 is substituted for A, and therefore provides an alternative to  
CC existing systems comprising a self-replicating HCV RNA molecule that, in  
CC conjunction with mutations in the HCV non-structural region, such as the  
CC G12042/C/R mutations, transduces and/or replicates with greater  
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
CC replicon Apk12 and contains the viral protease NS2/3, protease complex  
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
XX  
XX Sequence 2201 AA;  
SQ  
Query Match 86.4%; Score 1531; DB 5; Length 2201;  
Best Local Similarity 94.7%; Pred. No. 2,1e-139;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

OY	256	RRGDSRGSLSPRPVSYLKSSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET	315
Db	335	RRGDSRGSLSPRPVSYLKSSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET	394
OY	316	TMR	319
Db	395	TMR	398

Search completed: May 6, 2004, 09:30:43  
Job time : 46.463 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 10.8777 Seconds

(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771

Sequence: 1 MKKKLHHHHHTSAGITK.....TTWRTSSAMRHPRGKKKK 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1533	86.6	3010	1 A45573	genome polyprotein
2	1528	86.3	3010	1 GNMVCT	genome polyprotein
3	1515	85.5	3010	1 GNMVCT	genome polyprotein
4	1487	84.0	3010	1 S18030	genome polyprotein
5	1479	83.5	3010	1 GNMVCT	genome polyprotein
6	1404	79.3	3011	1 S40770	genome polyprotein
7	1399	79.0	3011	1 GNMVCT	genome polyprotein
8	1386	78.3	3011	1 GNMVCT	genome polyprotein
9	1235	69.7	3014	1 J05620	genome polyprotein
10	1173	66.2	3033	1 J01303	genome polyprotein
11	1158	65.4	3033	1 GNMVCT	genome polyprotein
12	398.5	22.5	3005	1 T08841	polyprotein - dour
13	343	19.4	2970	2 T08839	polyprotein - marm
14	108	6.1	353	2 G87392	DNA-directed DNA P
15	101	5.7	600	2 B46642	conserved hypochet
16	98.5	5.6	470	2 J04058	tetracycline 6-hyd
17	97.5	5.5	1085	2 T03531	coBN protein homol
18	95.5	5.4	706	2 S33761	transferin precur
19	94.5	5.3	716	2 G83612	hypothetical prote
20	93.5	5.3	7463	2 T36248	CD4 peptide synthe
21	93	5.3	660	2 VHMW2	structural protein
22	91	5.1	904	2 A84212	hypothetical prote
23	90.5	5.1	868	2 H81775	acetylcholinester
24	90	5.1	659	2 B44212	structural protein
25	90	5.1	2796	2 J04743	fatty-acid synthas
26	89.5	5.1	267	2 B83602	conserved hypochet
27	88	5.0	3414	1 GNMVCT	genome polyprotein
28	87	4.9	1057	2 T18171	hngl protein - hum
29	87	4.9	3412	1 GNMVCT	genome polyprotein

30	86.5	4.9	470	1 NMIVW8	exo-alpha-sialidas
31	86.5	4.9	990	2 S67499	glutamate synthase
32	85.5	4.8	299	2 AH3447	cytochrome-c oxida
33	85.5	4.8	348	2 H70549	probable pnh prot
34	85	4.8	470	1 NMIV9	exo-alpha-sialidas
35	85	4.8	707	2 D84154	cadmium-transport
36	84.5	4.8	347	2 S44167	malate dehydrogen
37	84.5	4.8	5627	2 C83339	hypothetical prote
38	84	4.7	223	2 T35594	hypothetical prote
39	84	4.7	3069	2 H70656	fatty-acid synthas
40	83.5	4.7	315	2 AG2361	hypothetical prote
41	83.5	4.7	538	2 S22409	D-alanyl-D-alanine
42	83.5	4.7	1399	2 G83112	DNA-directed RNA P
43	83	4.7	398	2 B71284	probable periplasm
44	83	4.7	2103	2 G86925	probable polyketid
45	83	4.7	4735	2 T17463	flamycin polyketi

## ALIGNMENTS

## RESULT 1

A45573 genome polyprotein - hepatitis C virus (strain J7)  
N/Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C/Accession: A45573  
R/Ianaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, I  
Virus Res. 23, 39-53, 1992  
A/Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s  
A/Reference number: A45573, PMID:9229574, PMID:1318627  
A/Accession: A45573  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-3010 <TRAN>  
A/Cross-references: GB:D11168; GB:D01171; NID:9221612; PID:BA001943.1; PID:9221613  
A/Experimental source: HCV-JT  
A/Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIPI:106207)  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywds: ATP; glycoprotein; hydroxase; nucleotide binding; P-loop; polyprotein; serin  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EPN>  
F/192-389/Product: major envelope protein E #status predicted <MEB>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepatitisvirin #status predicted <NS3>  
F/1130-1237/Region: nucleotide-binding motif A (P-loop)  
F/1132-1317/Region: nucleotide-binding motif B  
F/116-1862/Product: DEHX motif  
F/1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F/1663-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 86.6%; Score 1533; DB 1; Length 3010;

Best Local Similarity 95.1%; Pred. No. 2.7e-121; Matches 289; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY	16	AGITKPYEVRAGGLIRACGLVKAAGHYVQNAFKLALNTGYVYDHLPLQDVAHAG	75
DB	904	AATTAFTYFVRAGGLIRACGLVKAAGHYVQNAFKLALNTGYVYDHLPLQDVAHAG	963
QY	76	LRLAVAVEPVFSDMEVKITWAGDTAACGDIISGLPVSAARGREILIGPDNPEGQGM	135
DB	964	LRLAVAVEPVFSDMEVKITWAGDTAACGDIISGLPVSAARGREILIGPDNPEGQGM	1023
QY	136	RLAAPTAYSOQRLGLGCIITSITGRDNQVEGGEVQVSTATQSPFLATCNGVGVCTVH	195
DB	1024	RLAAPTAYSOQRLGLGCIITSITGRDNQVEGGEVQVSTATQSPFLATCNGVGVCTVH	1083
QY	196	GAGSKTLAPKPGITQMTYNNVDLWQAPPGARSMPTCTGSSDLYLVTRHADYIPVR	255

Db 1084 GAGSXTLAGPKGPIITOMYTNVDODLVGMHAPPGARSILPTCGSSDLYLTVTRHADYIPVR 1143  
 |||||  
 QY 256 RRGDSRGSILSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIPVESMET 315  
 |||||  
 Db 1144 RRGDSRGSILSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIPVESMET 1203  
 |||||  
 QY 316 TMRT 319  
 |||||  
 Db 1204 TMRS 1207

## RESULT 2

GNMVCJ  
 genome polyprotein - hepatitis C virus (strain J)  
 N/contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001  
 C/Accession: A39253; PS0086  
 R/Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto, Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
 A/Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
 A/Reference number: A39253; MUID:91088550; PMID:2175903  
 A/Accession: A39253

A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <KAT>  
 A/Cross-references: GB:D90208; NID:G221610; PIDN:BA14233.1; PID:G221611  
 R/Kato, N.; Ohkoshi, S.; Shimotohno, K. Proc. Jpn. Acad. 65B, 219-223, 1989  
 A/Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variability  
 A/Reference number: PS0086  
 A/Accession: PS0086  
 A/Molecule type: genomic RNA  
 A/Residues: 2650-2707 <KA2>  
 A/Experimental source: Japanese isolate

C/Comment: The cleavage sites of this polyprotein have not been determined.  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
 F/2-115/Product: capsid protein C #status predicted <CPC>  
 F/119-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <MEE>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1207-1237/Product: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Product: nucleotide-binding motif B  
 F/1316-1319/Product: DEXH motif  
 F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,223,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 86.3%; Score 1528; DB 1; Length 3010;  
 Best Local Similarity 94.1%; Pred. No. 7,1e-121;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVYDHLTPLODMAHAG 75  
 |||||  
 Db 904 AGITRVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVYDHLTPLODMAHAG 963  
 |||||  
 QY 76 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 135  
 |||||  
 Db 964 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 1023  
 |||||  
 QY 136 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGVVAVSTATOSFLATCNVGVCTVVF 195  
 |||||  
 Db 1024 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGVVAVSTATOSFLATCNVGVCTVVF 1083  
 |||||  
 QY 196 GAGSXTLAGPKGPIITOMYTNVDODLVGMHAPPGARSILPTCGSSDLYLTVTRHADYIPVR 255  
 |||||  
 Db 1084 GAGSXTLAGPKGPIITOMYTNVDODLVGMHAPPGARSILPTCGSSDLYLTVTRHADYIPVR 1143  
 |||||  
 QY 256 RRGDSRGSILSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIPVESMET 315

Db 1144 RRGDSRGSILSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIPVESMET 1203  
 |||||  
 QY 316 TMRT 319  
 |||||  
 Db 1204 TMRS 1207

## RESULT 3

GNMVCJ  
 genome polyprotein - hepatitis C virus (strain Taiwan)  
 N/contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 A/Note: host Homo sapiens (man)  
 C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C/Accession: A40244  
 R/Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
 Virology 189, 102-113, 1992  
 A/Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the

A/Reference number: A40244; MUID:92230206; PMID:1314449  
 A/Accession: A40244  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <CHB>  
 A/Cross-references: GB:M8754  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F/1-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <MEE>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1207-1237/Product: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Product: nucleotide-binding motif B  
 F/1316-1319/Product: DEXH motif  
 F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,223,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 85.5%; Score 1515; DB 1; Length 3010;  
 Best Local Similarity 92.8%; Pred. No. 9e-120;  
 Matches 282; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVYDHLTPLODMAHAG 75  
 |||||  
 Db 904 AGITRVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVYDHLTPLODMAHAG 963  
 |||||  
 QY 76 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 135  
 |||||  
 Db 964 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 1023  
 |||||  
 QY 136 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGVVAVSTATOSFLATCNVGVCTVVF 195  
 |||||  
 Db 1024 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGVVAVSTATOSFLATCNVGVCTVVF 1083  
 |||||  
 QY 196 GAGSXTLAGPKGPIITOMYTNVDODLVGMHAPPGARSILPTCGSSDLYLTVTRHADYIPVR 255  
 |||||  
 Db 1084 GAGSXTLAGPKGPIITOMYTNVDODLVGMHAPPGARSILPTCGSSDLYLTVTRHADYIPVR 1143  
 |||||  
 QY 256 RRGDSRGSILSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIPVESMET 315  
 |||||  
 Db 1144 RRGDSRGSILSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKADVPIPVESMET 1203  
 |||||  
 QY 316 TMRT 319  
 |||||  
 Db 1204 TMRS 1207

RESULT 4  
 S18030  
 genome polyprotein - hepatitis C virus (isolate JK1)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu

C:Species: hepatitis C virus

A:Variety: isolate JKI

C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001

C:Accession: S18030; S33570; A48332; S18029

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie

A:Reference number: S18028

A:Accession: S18030

A:Molecule type: genomic RNA

A:Residues: 1-3010 <HON>

A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479

A:Experimental source: isolate JKI from an individual

R:Honda, M.; Kaneko, S.; Uemura, M.; Kobayashi, K.; Murakami, S.

Arch. Virol. 148, 163-169, 1993

A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated.

A:Reference number: A48332; MUID:93119270; PMID:8380322

A:Accession: S33570

A:Molecule type: genomic RNA

A:Residues: 1-547, 'T', '548-621', 'V', '623-624', 'S', '626-652', 'DL', '655-761', 'T', '763-782' <HON>

A:Cross-references: EMBL:X61591

A:Note: this sequence is inconsistent with the nucleotide translation

A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320

as Trp, and TTC for residue 771 as Ser

A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:P121748)

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; p-loop; polyprotein; serin

F:12-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <ME>

F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>

F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEKH motif

F:1616-1862/Product: nonstructural protein NS4 #status predicted <N4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (as

Query Match 84.0%; Score 1487; DB 1; Length 3010;

Best Local Similarity 92.1%; Pred. No. 2.1e-117;

Matches 280; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 904 AGITRVPYFVRAOGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 963
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 135
Db 964 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 1023
QY 136 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSEFLATCNGVCWTVYH 195
Db 1024 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSEFLATCNGVCWTVYH 1083
QY 196 GAGSKTLAPKPGITOMTNTVODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVR 255
Db 1084 GAGSKTLAPKPGITOMTNTVODLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIPIVR 1143
QY 256 RRGDSRGSLSPPRPVSYLKGSGGPIILCPSGHANGIRAAVCTRGVAKAVDFIPVESMET 315
Db 1144 RRGDSRGSLSPPRPVSYLKGSGGPIILCPSGHANGIRAAVCTRGVAKAVDFIPVESMET 1203
QY 316 TMRT 319
Db 1204 TMRS 1207

```

RESULT 5

GENMTC genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu

C:Species: hepatitis C virus

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001

C:Accession: A38465

R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;

J. Virol. 65, 1105-1113, 1991

A:Title: Structure and organization of the hepatitis C virus genome isolated from human

A:Reference number: A38465; MUID:91140698; PMID:1847440

A:Accession: A38465

A:Molecule type: genomic RNA

A:Residues: 1-3010 <TAK>

A:Cross-references: EMBL:X58335; NID:G329770; PIDN:AAA72945.1; PID:G329771

A:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <ME>

F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>

F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEKH motif

F:1616-1862/Product: nonstructural protein NS4 #status predicted <N4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22,

Query Match 83.5%; Score 1479; DB 1; Length 3010;

Best Local Similarity 91.8%; Pred. No. 1e-116;

Matches 279; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 904 AGITRVPYFVRAOGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 963
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 135
Db 964 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 1023
QY 136 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSEFLATCNGVCWTVYH 195
Db 1024 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSEFLATCNGVCWTVYH 1083
QY 196 GAGSKTLAPKPGITOMTNTVODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVR 255
Db 1084 GAGSKTLAPKPGITOMTNTVODLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIPIVR 1143
QY 256 RRGDSRGSLSPPRPVSYLKGSGGPIILCPSGHANGIRAAVCTRGVAKAVDFIPVESMET 315
Db 1144 RRGDSRGSLSPPRPVSYLKGSGGPIILCPSGHANGIRAAVCTRGVAKAVDFIPVESMET 1203
QY 316 TMRT 319
Db 1204 TMRS 1207

```

RESULT 6

S40770 genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001

C:Accession: S40770; P01285

R:Okamoto, H.

submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A/Residues: 1-3011 <OKA>  
 A/Cross-references: EMBL:D10749; NID:G221586; PIDN:BA01582.1; PID:G221587  
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Ugn. J. Exp. Med. 60, 167-177, 1990  
 A/Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A/Reference number: PC1284; MUID:91013116; PMID:2170712  
 A/Accession: PC1285  
 A/Molecule type: genomic RNA  
 A/Residues: 1-513 <OK2>  
 A/Cross-references: GB:D00831; PID:G221511; PIDN:BA00705.1; PID:G221512  
 A/Experimental source: isolate HC-U1  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F/2-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <ME>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 Query Match 79.3%; Score 1404; DB 1; Length 3011;  
 Best Local Similarity 84.5%; Pred. No. 2.4e-110;  
 Matches 257; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

QY 16 AGITKVPFVFAOGILRACMLVRKAAGHYQVAFMCLALTGTGVYDHLTPLODMAHG 75  
 DB 904 ASLKVPEFVRVQGLRICALARKVGVGHVQVMIITLGAITGVYVNHLPDWMANG 963  
 QY 76 LRDVAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARRGRIILGPADNFBGQW 135  
 DB 964 LRDVAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARRGRIILGPADNFBGQW 1023  
 QY 136 RLAPITAYSOOTRELGLCIITSLTGPRDKNOVEGVQVVSATOSFLATCVNGVCTVFEH 195  
 DB 1024 RLAPITAYSOOTRELGLCIITSLTGPRDKNOVEGVQVVSATOSFLATCVNGVCTVFEH 1083  
 QY 196 GAGSKTLGAPKPIQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 255  
 DB 1084 GAGRTTISPKGPVQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 1143  
 QY 256 RRGDSRGSILSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSILSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 7  
 GNMVCH  
 genome polyprotein - hepatitis C virus (strain HCV-1)  
 N/contains: capsid protein C; envelope protein M; hepatitis A (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/species: hepatitis C virus  
 C/date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 19-Jan-2001  
 C/Accession: A39166; PQ0403; PQ0404  
 R/Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
 A/Title: Genetic organization and diversity of the hepatitis C virus.  
 A/Reference number: A39166; MUID:91172826; PMID:1848704  
 A/Accession: A39166  
 A/Molecule type: mRNA  
 A/Residues: 1-3011 <COC>  
 A/Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874  
 R/Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L  
 J. Gen. Virol. 73, 1131-1141, 1992

A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
 A/Reference number: PQ0393; MUID:92268871; PMID:1316939  
 A/Accession: PQ0403  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CHA>  
 A/Cross-references: DBJ:D10128  
 A/Experimental source: isolates E-b16  
 A/Accession: PQ0404  
 A/Status: preliminary  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CH2>  
 A/Experimental source: isolates E-b17  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
 F/1-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <ME>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,222  
 Query Match 79.0%; Score 1399; DB 1; Length 3011;  
 Best Local Similarity 84.2%; Pred. No. 6.3e-110;  
 Matches 256; Conservative 25; Mismatches 23; Indels 0; Gaps 0;

QY 16 AGITKVPFVFAOGILRACMLVRKAAGHYQVAFMCLALTGTGVYDHLTPLODMAHG 75  
 DB 904 ASLKVPEFVRVQGLRICALARKVGVGHVQVMIITLGAITGVYVNHLPDWMANG 963  
 QY 76 LRDVAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARRGRIILGPADNFBGQW 135  
 DB 964 LRDVAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARRGRIILGPADNFBGQW 1023  
 QY 136 RLAPITAYSOOTRELGLCIITSLTGPRDKNOVEGVQVVSATOSFLATCVNGVCTVFEH 195  
 DB 1024 RLAPITAYSOOTRELGLCIITSLTGPRDKNOVEGVQVVSATOSFLATCVNGVCTVFEH 1083  
 QY 196 GAGSKTLGAPKPIQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 255  
 DB 1084 GAGRTTISPKGPVQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 1143  
 QY 256 RRGDSRGSILSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSILSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 8  
 GNMVCH  
 genome polyprotein - hepatitis C virus (strain H)  
 N/contains: capsid protein C; envelope protein M; hepatitis A (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/species: hepatitis C virus  
 A/Note: host Homo sapiens (man)  
 C/date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C/Accession: A36814; A41546  
 R/inchausti, G.; Zepede, S.; Lee, D.H.; Sugtani, M.; Nasoff, M.; Prince, A.M.  
 Submitted to GenBank, July 1992  
 A/Description: Genomic structure of the human prototype strain H of hepatitis C virus.  
 A/Reference number: A36814  
 A/Accession: A36814  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3011 <INC>



A/Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
 Rinchanspe, G.; Zebadee, S.; Lee, D.H.; Sugita, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A/Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison  
 A/Reference number: AA1546; MUID:92052256; PMID:1658800  
 A/Contents: annotation  
 A/Note: neither amino acid nor nucleotide sequence is given  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein C #status predicted <CPC>  
 F.1115/Product: capsid protein C #status predicted <CPC>  
 F.116-191/Product: envelope protein M #status predicted <EPM>  
 F.192-389/Product: major envelope protein E #status predicted <ME>  
 F.730-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F.1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F.1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F.1312-1317/Region: nucleotide-binding motif B  
 F.1316-1319/Region: DEXH motif  
 F.1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F.1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F.2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F.196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 78.3%; Score 1386; DB 1; Length 3011;  
 Best Local Similarity 83.6%; Pred. No. 8e-109;  
 Matches 254; Conservative 27; Mismatches 23; Indels 0; Gaps 0;

QY 16 AGITVFYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 75  
 Db 904 ASLKVYFVFAVQGLIKALIKAKIGHYVMAITGLTGVNHLAPLDMAHAG 963

QY 76 LRDAVAVEPVIQSDMEVKIITWGAADTACGDIISGLPVSARRREITLGPADNFEQGW 135  
 Db 964 LRDAVAVEPVIQSDMEVKIITWGAADTACGDIISGLPVSARRREITLGPADNFEQGW 1023

QY 136 RLAPITAYSGOQTGLIGCIITSLTGRDKNOVEGVVSTATOSPLATCNGVCMVFEH 135  
 Db 1024 RLAPITAYSGOQTGLIGCIITSLTGRDKNOVEGVVSTATOSPLATCNGVCMVFEH 1083

QY 196 GAGSKTLAGEKPTITQWYTNVDODLVGMOAPPGASMTPTCCSSDLYLVRHADYIPVR 255  
 Db 1084 GAGSKTLAGEKPTITQWYTNVDODLVGMOAPPGASMTPTCCSSDLYLVRHADYIPVR 1143

QY 256 RRGDSRGLSPRVSTLXKSGGGLCPGSHANGITRAVCTRGVAKADVIFVESMET 315  
 Db 1144 RRGDSRGLSPRVSTLXKSGGGLCPGSHANGITRAVCTRGVAKADVIFVESMET 1203

QY 316 TMRT 319  
 Db 1204 TMRT 1207

RESULT 9  
 JC5620 genome polyprotein - hepatitis C virus (isolate EUH1480)  
 N/Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C/Accession: J05620  
 R/Chamblain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A/Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
 A/Reference number: J05620; MUID:9736593; PMID:9223423  
 A/Accession: J05620  
 A/Molecule type: mRNA  
 A/Residues: 1-3014 <CHA>  
 A/Cross-references: GB:Y13184  
 A/Experimental source: genotype 5a, which predominates in South Africa  
 A/Note: the translation of the nucleotide sequence is not complete in this paper  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F.2-115/Product: capsid protein C #status predicted <CPC>

F.116-191/Product: envelope protein M #status predicted <EPM>  
 F.192-389/Product: major envelope protein E #status predicted <ME>  
 F.734-408/Region: hypervariable #status predicted <ME>  
 F.730-730/Product: nonstructural protein NS1 #status predicted <NS1>  
 F.731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F.1008-1616/Product: nonstructural protein NS2 #status predicted <NS2>  
 F.1231-1238/Region: nucleotide-binding motif A (P-loop)  
 F.1313-1318/Region: nucleotide-binding motif B  
 F.1317-1320/Region: DEXH motif  
 F.1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F.1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F.2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
 F.2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 69.7%; Score 1235; DB 1; Length 3014;  
 Best Local Similarity 72.5%; Pred. No. 5.3e-96;  
 Matches 219; Conservative 44; Mismatches 39; Indels 0; Gaps 0;

QY 18 ITXVFYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 77  
 Db 907 LTXVFYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 966

QY 78 DLAVAVEPVIQSDMEVKIITWGAADTACGDIISGLPVSARRREITLGPADNFEQGW 137  
 Db 967 ELTVAVEPVIQSDMEVKIITWGAADTACGDIISGLPVSARRREITLGPADNFEQGW 1026

QY 138 LAPITAYSGOQTGLIGCIITSLTGRDKNOVEGVVSTATOSPLATCNGVCMVFEH 197  
 Db 1027 LAPITAYSGOQTGLIGCIITSLTGRDKNOVEGVVSTATOSPLATCNGVCMVFEH 1086

QY 198 GSKTLAGEKPTITQWYTNVDODLVGMOAPPGASMTPTCCSSDLYLVRHADYIPVR 257  
 Db 1087 GSKTLAGEKPTITQWYTNVDODLVGMOAPPGASMTPTCCSSDLYLVRHADYIPVR 1146

QY 258 RRGDSRGLSPRVSTLXKSGGGLCPGSHANGITRAVCTRGVAKADVIFVESMET 317  
 Db 1147 RRGDSRGLSPRVSTLXKSGGGLCPGSHANGITRAVCTRGVAKADVIFVESMET 1206

QY 318 RT 319  
 Db 1207 RS 1208

RESULT 10  
 J01303 genome polyprotein - hepatitis C virus (isolate HC-06)  
 N/Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C/Accession: J01303  
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Kura, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;  
 J. Gen. Virol. 72, 2697-2704, 1991  
 A/Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum  
 A/Reference number: J01303; MUID:92044440; PMID:165196  
 A/Accession: J01303  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3033 <OKA>  
 A/Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651  
 A/Experimental source: isolate HC-06 from a Japanese individual  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transme  
 F.2-115/Product: capsid protein C #status predicted <CPC>  
 F.116-191/Product: envelope protein M #status predicted <EPM>  
 F.192-389/Product: major envelope protein E #status predicted <ME>  
 F.734-1010/Product: nonstructural protein NS1 #status predicted <NS1>  
 F.1011-1619/Product: nonstructural protein NS2 #status predicted <NS2>  
 F.1316-1321/Region: nucleotide-binding motif A (P-loop)  
 F.1318-1323/Region: DEXH motif  
 F.1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F.1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F.2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 66.2%; Score 1173; DB 1; Length 3033;  
Best Local Similarity 69.2%; Pred. No. 9,8e-91;  
Matches 209; Conservative 45; Mismatches 48; Indels 0; Gaps 0;

18 ITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTFLQDMAHAGIR 77

910 IIRVPEFVRAHALLRRCNTVRHLAAGRYVQWVLLALGRMTGYIYDHLTPMSDMAANGIR 969

78 DLAAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 137

970 DLAAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 1029

138 LAITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 197

1030 LAITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 1089

198 GSXTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 257

1090 GSKTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 1149

258 GDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 317

1150 GDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 1209

318 RT 319

1210 RS 1211

# RESULT 11

GNMWV8

Genome polyprotein - hepatitis C virus (strain HC-J8)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (HCV) genome having poor homology to rep

protein NS4; nonstructural protein NS5b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 19-Jan-2001

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

C:Accession: A40250; F00397; F00359

F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 65.4%; Score 1158; DB 1; Length 3033;  
Best Local Similarity 67.3%; Pred. No. 1.8e-89;  
Matches 206; Conservative 47; Mismatches 53; Indels 0; Gaps 0;

16 AGITVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTFLQDMAHAG 75

908 ASLIRIFVRAHALLRRCNTVRHLAAGRYVQWVLLALGRMTGYIYDHLTPMSDMAANG 967

76 LRLDAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 135

968 LRLDAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 1027

136 RLAPITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 195

1028 RLAPITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 1087

196 GAGSXTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 255

1088 GAGSXTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 1147

256 RRGDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 315

1148 RRGDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 1207

316 TMRTSS 321

1208 ATRTSS 1213

# RESULT 12

TO8841

polyprotein - douroucouli hepatitis GB virus A

C:Species: douroucouli hepatitis GB virus A

C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text\_change 17-Nov-2000

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 65.4%; Score 1158; DB 1; Length 3033;  
Best Local Similarity 67.3%; Pred. No. 1.8e-89;  
Matches 206; Conservative 47; Mismatches 53; Indels 0; Gaps 0;

16 AGITVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTFLQDMAHAG 75

908 ASLIRIFVRAHALLRRCNTVRHLAAGRYVQWVLLALGRMTGYIYDHLTPMSDMAANG 967

76 LRLDAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 135

968 LRLDAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 1027

136 RLAPITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 195

1028 RLAPITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 1087

196 GAGSXTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 255

1088 GAGSXTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 1147

256 RRGDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 315

1148 RRGDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 1207

316 TMRTSS 321

1208 ATRTSS 1213

# RESULT 12

TO8841

polyprotein - douroucouli hepatitis GB virus A

C:Species: douroucouli hepatitis GB virus A

C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text\_change 17-Nov-2000

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 65.4%; Score 1158; DB 1; Length 3033;  
Best Local Similarity 67.3%; Pred. No. 1.8e-89;  
Matches 206; Conservative 47; Mismatches 53; Indels 0; Gaps 0;

16 AGITVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTFLQDMAHAG 75

908 ASLIRIFVRAHALLRRCNTVRHLAAGRYVQWVLLALGRMTGYIYDHLTPMSDMAANG 967

76 LRLDAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 135

968 LRLDAVEEPIFSDMEVKIITWGAADTAACGDIISGLFVSARRREILLGADNPEQGRIL 1027

136 RLAPITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 195

1028 RLAPITVYSGOQTRGLGCIITSLTGRDKQVGEVQVSTATOSFLATCVNGVCMVTFH 1087

196 GAGSXTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 255

1088 GAGSXTLAGPKPITOMYTNVDQVLGWQAPPGARSMTPTCGSSDLVLYTRHADVIPIVR 1147

256 RRGDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 315

1148 RRGDSRGLSPRPVSYLKGSGGGLPCPSGHAAGIFRAAAGTGRVAKAVDFIPVESMET 1207

316 TMRTSS 321

1208 ATRTSS 1213

# RESULT 12

TO8841

polyprotein - douroucouli hepatitis GB virus A

C:Species: douroucouli hepatitis GB virus A

C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text\_change 17-Nov-2000

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

C:Accession: TO8841

QY 287 HAVGIFRAVCTRGVAKAVDFIPVESMTTMTS 320  
 Db 1124 HVGGMV-VSVLARGV-KATGVRVYKPKMETLTKDS 1155

## RESULT 13

108839  
 POLYPROTEIN - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: 108839  
 R/Enter: U.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: 216486; PMID:98120818; PMID:9460920  
 A/Accession: 108839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <SRK>  
 A/Cross-references: EMBL:AF023424; NID:92828597; PIDN:AA040501.1; PID:92828598  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 19.4%; Score 343; DB 2; Length 2970;  
 Best Local Similarity 30.5%; Pred. No. 2, 7e-20;

Matches 85; Conservative 44; Mismatches 94; Indels 56; Gaps 9;

QY 71 WAHAG-----LRDLAAVEPVIFSDMEVKITMGADTAACGDIISGLPVSARRGRE 121  
 Db 891 YAHAGVTRTAEDELQMGFALFVAVHPEDCAMVRAATLSCGSGVHGKPVARRGDE 950  
 QY 122 ILGPADNFEQGMRL-----LAPITAYSQQTGILGCIITSLTGKDKQVGEVQVVS 175  
 Db 951 VLIGVLNGV---WELPPGFVPTAPVAVH-HHGKFGGVKTSWTDTEHGVNVVVG 1005  
 QY 176 TATGSPFLATGVNGVCMVFHAGSKTLGAPKGPITQMTYVDDLVGMQAPPGARSMTPC 225  
 Db 1006 TSTIRSGTCVNGVMTYTHGSSNARLTAAQMGPNRMSASDVAVYPLPFAKCLEEP 1065  
 QY 236 TCGSSDLYLVTRHADVIFVRRGDSRGLS-----PRPVSYLKSSSGEPILCP 284  
 Db 1066 KCGQGVWV-----RND--GALCHGTIGRTVELDLAELCDPFGSSGSFILCD 1112  
 QY 285 SGHAGVIFRAVCTRGVAKAVDFIPVESMTTMTSSAW 323  
 Db 1113 EGHAVGML-LSVLRG-----SRVTGIRYTKFW 1139

## RESULT 14

887392  
 conserved hypothetical protein CC1155 [imported] - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: G87392  
 R/Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.S.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjol  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; PMID:1173698; PMID:11259647  
 A/Accession: G87392  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1353 <STO>  
 A/Cross-references: GB:AE005673; NID:G13422473; PIDN:AAK23139.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CC1155

Query Match 6.1%; Score 108; DB 2; Length 353;  
 Best Local Similarity 21.9%; Pred. No. 0.16;  
 Matches 77; Conservative 36; Mismatches 122; Indels 116; Gaps 14;

QY 22 PYFVRAQGLIRACMLVEXAA-----GHHV-----QMAFMKLAALNGTYVDHL 65  
 Db 65 PIVVLAGLPAISQALRESAIVAMRASGISGRITGMVPAVAVVLLDALCGVLAAPRA 124  
 QY 66 TP-LQDWAHAGLDLAVAEPIVIFSDMEVKITMGADTAACGDIISGLPVSARRGREILL 124  
 Db 125 DPLTADM-WRNTTPVAREKEPVPRTFRAGADIVIGNANASDRITITGVITFRDSGLIV 163  
 QY 125 ----GPADNFEQGMRLAPITAYSQQTGILGCIITSLTGKDKQVGEVQVVSATOS 180  
 Db 184 EKVAPAAVYDKAMVLEQPKT-----TRFADLSQASTPA-- 219  
 QY 181 FLATGVNGVCMVFHAGSKTLGAPKGPITQMTYVDDLVGMQAPPGARSMTPTCGSS 240  
 Db 220 ----ATSWP-----TALRPQDVGLFDDDSVFLTAAS----- 246  
 QY 241 DLYLVTRHADVIFVRRGDSRGLSPRPVSY---LKGSSGGP-----ILCFSGHAYG 290  
 Db 247 -----ARRALENGG---SDRPESFYATHLQAFASPFVSLVWLLLSAPVALA 290  
 QY 291 IFR---AAVCTRGVAKAVDFIPVESMTTMTSSA-----WRHPQCG 330  
 Db 291 NFRSGGAVLLTGLAGCMFLVANGMLTALDESGALTPFLAVMAFAIFG 341

## RESULT 15

B46642  
 DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68k chair  
 C/Species: Mus musculus (house mouse)  
 C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
 C/Accession: B46642  
 R/Miyazawa, H.; Izumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.  
 J. Biol. Chem. 268, 8111-8122, 1993  
 A/Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase  $\alpha$   
 A/Reference number: A46642; PMID:93216788; PMID:8463324  
 A/Accession: B46642  
 A/Status: preliminary  
 A/Molecule type: mRNA; protein  
 A/Residues: 1-600 <MTY>  
 A/Cross-references: GB:DL3546; NID:9303658; PIDN:BA02746.1; PID:9303659  
 A/Experimental source: FMA3 cells  
 A/Note: sequence extracted from NCBI backbone (NCBIN:129148, NCBI:129149)  
 C/Keywords: nucleic acidyltransferase

Query Match 5.7%; Score 101; DB 2; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 1.3;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 30 LIRACMLVRKKAAGHYVGM-AFMKLAALT-----GTYVDHL-----TFLQWMA 72  
 Db 27 LAELCVLRQTEDEGWSVSLIAFCTISAGTCTLVDIANSFEYEVNKKLSKAWHSASKDSG 86  
 QY 73 HAGLDLAVAEPIVIFSDMEVKITMGADTAACGDI--ISGLP-----VSARRGREI 122  
 Db 87 HAGTNDI-VSIQELIAREEBEETLLSTTSKGLKRVSSSTPEPLTKRVAARSPPQ- 144  
 QY 123 ILGPADNFEQGMRLAPITAYSQQTGILGCIITSLTGKDKQVGEVQVVSATOSFL 182  
 Db 145 LLSPPSS-----FSPSATPSPQK-----YTSRTNR-----GEVVTTFGSAQ--- 178  
 QY 183 ATCVNGVCMVFHAGSKTL--AGPKGPITQMTYVDDLVG 222  
 Db 179 ----GLSMGRGSGSVSLKVVDPEPLTGSYRAMFQQLNG 215

Search completed: May 6, 2004, 09:37:14  
 Job time : 12.8777 secs



Fri May 7 13:37:08 2004

us-10-650-585-4.rsp

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.97286 Seconds

(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-4

Perfect score: 1771

Sequence: 1 MKKKKLEHHHHHTSAGITK.....TTWRTSSAWHPQFGKKKK 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1533	86.6	3010	POLG_HCVUT	Q00269 h genome po
2	1528	86.3	3010	POLG_HCVJA	P26662 h genome po
3	1515	85.5	3010	POLG_HCVTA	P29846 h genome po
4	1479	83.5	3010	POLG_HCVBK	P26663 h genome po
5	1399	79.0	3011	POLG_HCVL	P26664 h genome po
6	1386	78.3	3011	POLG_HCVH	P27948 h genome po
7	1173	66.2	3033	POLG_HCVU6	P26660 h genome po
8	1158	65.4	3033	POLG_HCVU8	P26661 h genome po
9	101	5.7	600	P202_HORSE	P33611 mus musculi
10	95.5	5.4	706	TRFE_HORSE	P27425 equus caball
11	93	5.3	660	VST2_HEYBU	P33426 hepatitis e
12	93	5.3	660	VST2_HEYPA	P70289 mus musculi
13	91.5	5.2	1705	PTPV_MOUSE	Q03500 hepatitis e
14	90	5.1	659	VST2_HEYME	P14336 t genome po
15	88	5.0	3414	POLG_TBEMV	P07720 t genome po
16	87	4.9	3412	POLG_TBEMV	P00129 t genome po
17	87	4.9	3414	POLG_TBEMV	P05803 influenza a
18	86.5	4.9	470	NRAM_IAMHM	O8860 chlorobium
19	86	4.9	434	TOLB_CHLTE	P03472 influenza a
20	85	4.8	470	NRAM_IATRA	O8867 methanosa
21	85	4.8	730	HELS_METWA	P46487 methanosa
22	84.5	4.8	347	MDHM_EUCGU	P56500 rattus norv
23	84	4.7	309	UCP2_PAT	O8190 coxynebac
24	84	4.7	339	UCP2_PAT	P1510 influenza a
25	84	4.7	470	NRAM_IARUE	P39045 actinodac
26	83.5	4.7	538	DAC_ACTSP	O9C1X1 schizosac
27	83.5	4.7	854	FMP2_SCHPO	O9C1X1 schizosac
28	83.5	4.7	1399	RPOC_PSEAE	O9HWC9 pseudomon
29	83	4.7	309	UCP2_HUVA	P55851 homo sapien
30	83	4.7	485	VST2_HEYVA	O00270 hepatitis e
31	83	4.7	485	VST2_HEYVA	O04611 hepatitis e
32	82.5	4.7	453	NRAM_IAMIL	P03470 influenza a
33	82	4.6	309	UCP2_MOUSE	P70406 mus musculi

#### ALIGNMENTS

```
RESULT 1
ID POLG_HCVUT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-UT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitisvirus.
CX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals."
RL Virus Res. 23:39-53(1992).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position. Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
CC EMBL, D11168; BAA01943.1; -.
CC PIR, A45573; A45573.
CC DR PIR, A45573; A45573.
CC DR MEROPS, S29.001; -.
CC DR MEROPS, U39.001; -.
CC DR InterPro, IPR009003; Cys_Ser_trypsin.
CC DR InterPro, IPR001410; DEAD.
CC DR InterPro, IPR002522; HCV_capsid.
```

DR InterPro: IPR002552; HCV\_core.  
 DR InterPro: IPR002531; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5.  
 DR InterPro: IPR002166; HCV\_RDR.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVLr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RDR; 1.  
 DR Pfam: PF00487; DEXDC; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 DR Transmembrane; Nonstructural protein; Coat protein; Envelope protein C BY THE  
 DR INT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT CHAIN 3010 369  
 FT TRANSMEM 347 1083  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2529 2529  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C7435D642B CRC64;  
 Query Match 86.6%; Score 1533; DB 1; Length 3010;  
 Best Local Similarity 95.1%; Pred. No. 9,9e-124;  
 Matches 289; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 504 AAITAMPYFVRAQGLIRACMLVRKAVAGHYVQAFMKLAALTGTYYVDHLTPIDQNAHAG 963  
 QY 76 LRDIAVAEPIFSDMEVKIITMGADTAACGDIISGLPVARRGREIILGPADNFEQGV 135  
 Db 964 LRDIAVAEPIFSDMEVKIITMGADTAACGDIISGLPVARRGREIILGPADNFEQGV 1023  
 QY 136 RLAPITAYSOQTGGLGCIITSLTGRDKQVEGEVQVSTATISFLATCVNGYCVTFH 195  
 Db 1024 RLAPITAYSOQTGGLGCIITSLTGRDKQVEGEVQVSTATISFLATCVNGYCVTFH 1083  
 QY 196 GAGSKTLAAGPKGPTTQMTWNTVDOLVGMQAPPGARSTPTCGSSDLVLTFRADVTPVR 255  
 Db 1084 GAGSKTLAAGPKGPTTQMTWNTVDOLVGMQAPPGARSTPTCGSSDLVLTFRADVTPVR 1143  
 QY 256 RRGDSRGSLSPREVSYLKGSSGPTLCPGSHAVGIFRAAVCTRGVAKAVDFIPESMET 315  
 Db 1144 RRGDSRGSLSPREVSYLKGSSGPTLCPGSHAVGIFRAAVCTRGVAKAVDFIPESMET 1203  
 QY 316 TWRT 315  
 Db 1204 TWRS 1207  
 RESULT 2  
 ID POLG\_HCVUA STANDARD; PRT; 3010 AA.  
 AC P26662;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [containing: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP1) (GP15); Envelope glycoprotein E2  
 DE (GP68) (GP10) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Proteinase/helicase NS3 (P70) (Hepaticin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Japanese) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91088550; PubMed=2175903;  
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.;  
 RT "Molecular cloning of the human hepatitis C virus genome from  
 RT Japanese patients with non-A, non-B hepatitis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
 RN [2]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=91192160; PubMed=1849488;  
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,  
 RA Ohkoshi S., Shimotohno K.;  
 RT "Molecular structure of the Japanese hepatitis C viral genome";  
 RL FEBS Lett. 280:325-328(1991).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
 CC {RNA} (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its





DR EMBL; M84754; -; NOT\_ANNOTATED\_CDS.

DR  
MEROPS;

DR Interpretation

DR  
InterPro

DR Interpret

DR  
Interp

DR P. I. A. M.; P. I. A. M.

DR FLAM, E  
DR DFAM, P

DR Fall, F  
DR DFam, D

DR  
EMADT,  
E100011,

3D-STM  
KIM

CHATTIN

FT  
CHAIN

ET CHAIN

FT ACT SIT

FT SITE

ET CARBOHY

FT CARBOHY

ET CARBOHY

•

Q. 3. Mr. + Dr.

?

8

22 30

22

1000

222

POLG\_HU

10-001-

25 (E) 3.4

US  
NEPALIC  
05

RECEIVED  
JUN 11 1964

01438147  
OILSILT  
KA  
DE

RECEIVED  
JUL 27 1967

DOLORES  
"NON-ET  
PT





```

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MM; 658C9447CE5AF9 CR664;

Query Match 79.0%; Score 1399; DB 1; Length 3011;
Best Local Similarity 84.2%; Pred. No. 3.6e-112;
Matches 256; Conservative 25; Mismatches 23; Indels 0; Gaps 0;

16 AGITKVPYFVNAAGLIRACMLVRKAAGHYVMAFMKLAALTGYVDHLPLDPAHAG 75
Db 904 ASLKVLPFVAVQGLRFCALARKMIGHYVMYIKLGALTGYVNHLPFLDMAHNG 963
Qy 76 LRDLAVAVEPYIFSDMEYKILTWGADTAACGDIISGLPVSAARRREILLGPADNFEQGW 135
Db 964 LRDLAVAVEPYIFSDMEYKILTWGADTAACGDIISGLPVSAARRREILLGPADNFEQGW 1023
Qy 136 RLAPITAYSOOTKGLGICITSLTGRDNQVBEQVYVSTATOSPLATCYNGVQWTFH 195
Db 1024 RLAPITAYSOOTKGLGICITSLTGRDNQVBEQVYVSTATOSPLATCYNGVQWTFH 1083
Qy 196 GAGSKTIAGPSPITOMYNTVODLVGMQAPRGARSMPTCCGSSDLYLVTRHADVPVR 255
Db 1084 GAGSKTIAGPSPITOMYNTVODLVGMQAPRGARSMPTCCGSSDLYLVTRHADVPVR 1143
Qy 256 RRGDSRGSLSPRPVSYIKSSGGPDLCPGSHAVGIFPAAVCTRGVAKAVDFIPESMET 315
Db 1144 RRGDSRGSLSPRPVSYIKSSGGPDLCPGSHAVGIFPAAVCTRGVAKAVDFIPESMET 1203
Qy 316 TMRT 319
Db 1204 TMRT 1207

RESULT 6
POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P67); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P67) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

```

```

EX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cadle M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467 (1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=96154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100 (1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67463; AAA4534.1; -
DR PIR; A36814; GNVACH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AUR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR INTERPRO; IPRO090003; Cys_Ser_trypsin.
DR INTERPRO; IPRO01410; DEAD.
DR INTERPRO; IPRO02522; HCV capsid.
DR INTERPRO; IPRO02521; HCV core.
DR INTERPRO; IPRO02519; HCV env.
DR INTERPRO; IPRO02518; HCV NS1.
DR INTERPRO; IPRO02517; HCV NS2.
DR INTERPRO; IPRO00745; HCV NS4A.
DR INTERPRO; IPRO01490; HCV NS4B.
DR INTERPRO; IPRO02868; HCV NS5A.
DR INTERPRO; IPRO02166; HCV RdRp.
DR INTERPRO; IPRO01650; Helicase_C.
DR INTERPRO; IPRO04109; Peptidase_C29.
DR INTERPRO; IPRO07095; RNA_pol_D5_PS.
DR INTERPRO; IPRO07094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.

```





CC HepaC:Virus.  
 CC NCBI\_TaxID=11113;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=92044440; PubMed=1658196;  
 CC Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,  
 CC Machida A., Miyakawa Y., Mayumi M.,  
 CC "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 CC from a human carrier: comparison with reported isolates for conserved  
 CC RT and divergent regions.";  
 CC J. Gen. Virol. 72:2697-2704 (1991).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; D00944; BA00792.1; -;  
 DR PIR; J01303; J01303.  
 DR HSSP; P27958; 1HE1.  
 DR MEROPS; S29.001; -;  
 DR MEROPS; U39.001; -;  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4A.  
 DR InterPro; IPR001490; HCV\_NS4B.  
 DR InterPro; IPR002868; HCV\_NS5A.  
 DR InterPro; IPR002166; HCV\_NS5B.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C9.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4A; 1.  
 DR Pfam; PF01001; HCV\_NS4B; 1.  
 DR Pfam; PF01506; HCV\_NS5A; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 CC Transmembrane; Nonstructural  
 CC protein; Hydrolyase; Serine protease.  
 CC REMOVED FROM CAPSID PROTEIN C BY THE  
 CC CELLULAR AMINOPEPTIDASE.  
 CC CAPSID PROTEIN C (POTENTIAL).  
 CC MATRIX PROTEIN (POTENTIAL).

FT CHAIN 192 363 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).  
 FT CHAIN 734 1010 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT TRANSMEM 347 369 POTENTIAL.  
 FT ACT SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP BIND 1234 1241 ATP (POTENTIAL).  
 FT SITE 1320 1332 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SEQUENCE 3033 AA; 329165 MW; P957F5C1A273B5B5 CRC64;

Query Match 66.2%; Score 1173; DB 1; Length 3033;  
 Best Local Similarity 69.2%; Pred. No. 1.2e-92;  
 Matches 209; Conservative 45; Mismatches 48; Indels 0; Gaps 0;

QY 18 ITKVPYFVARAGLIRPCMKVRKAGSHYOMFMKLAALGTGVYVDHLPLDMMHAGGR 77  
 DB 910 LTRVPTFVRHMLALNRCTVRLAGRYOMVLLAGRTGTYI1DHLTMSDMANGRR 969  
 QY 78 DLAVAVEPIVFSDFMEKIIITWGDATACGDIISGLVPSARRREIILGPADEEGGML 137  
 DB 970 DLAVAVEPIVFSDFMEKVIWVGAEVTAACGDIILGVPVSRRLGREVLLGPADEGYSKGMEL 1029  
 QY 138 LAPITVYSOOTRGLGICITSLTGRDKNOVEBEVQVVSATATSPFLATCNVGVWYFHGA 197  
 DB 1030 LAPITVYAOQTRGLGICITVSMGRDKTEQABEIOVLSVTYQSFGTTISGVLMVYHGA 1089  
 QY 198 GSKTLAGPKGPIITOMVTNVODLVGWOAPPGARSMPTCCGSSDYLVTYRHADVIFVRR 257  
 DB 1090 GSKTLAGSGPVTOMVSSAEGDLVGMPSPPGKSLSEPCCGAVDLYLVTRNMDVIFARR 1149  
 QY 258 GDSRGSLLSPRPVSYSIKSGSGGELLCPSGHGVGIFRAAVCTGVAKADVIFVESMETTM 317  
 DB 1150 GDSRGSLLSPRPVSYSIKSGSGGELLCPSGHGVGIFRAAVCTGVAKADVIFVESMETTM 317  
 QY 318 RT 319  
 DB 1210 RS 1211

RESULT 8  
 POLG\_HCV8 STANDARD; PRT; 3033 AA.  
 ID POLG\_HCV8  
 AC P2661;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (P34.4.22.-); Protease/helicase NS3 (P70) (Hepatitis C virus)].

(EC 3.4.21.98): Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48) 1.  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; HCV.  
 NCBI\_TaxID=11115;  
 [1]  
 SEQUENCE FROM N.A. PubMed=1314459;  
 MEDLINE=92230232;  
 Okamoto H., Kuzai K., Okada S.-I., Yamamoto K., Iizuka H., Tanaka T., Fukuda S., Tsuda F., Mishiro S.;  
 RA "Full-length sequence of a hepatitis C virus genome having poor RT  
 homology to reported isolates: comparative study of four distinct genotypes";  
 RT Virology 188:331-341 (1992).  
 CC  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and RNA.  
 CC  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D10988; EAA01761.1; -  
 DR PIR: A40250; GNMVJB.  
 DR HSSP: P27958; 1HEI.  
 DR MEROPS: S29.001; -  
 DR MEROPS: U39.001; -  
 DR InterPro: IPR009003; Cys\_Ser\_Trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4A.  
 DR InterPro: IPR001490; HCV\_NS4B.  
 DR InterPro: IPR002868; HCV\_NS5A.  
 DR InterPro: IPR002166; HCV\_NS5B.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4A; 1.  
 DR Pfam: PF01001; HCV\_NS4B; 1.  
 DR Pfam: PF01506; HCV\_NS5A; 1.  
 DR Pfam: PF00998; Viral\_RDRP; 1.  
 DR Pfam: PD186062; HCV\_NS1; 1.  
 DR SMART: SMO0487; DEXDC1; 1.  
 KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KM Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1  
 REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 115  
 FT CHAIN 116 181  
 FT CHAIN 192 393  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT ACT\_SITE 1234 1241  
 FT NP\_BIND 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 299 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 FT CARBOHYD 3033 3033  
 FT SEQUENCE 330177 MW; 1A173E7E3381FD1A CRC64;  
 Query Match 65.4%; Score 1158; DB 1; Length 3033;  
 Best Local Similarity 67.3%; Pred. No. 2,3e-91;  
 Matches 206; Conservative 47; Mismatches 53; Indels 0; Gaps 0;  
 QY 16 AGITVVPYFVBAQGLIRACMLVRKAGGYQMAFMKLAALTGYVVDHLPLODMANAG 75  
 DB 908 ASLIRPIFYRAHALRCTLVHLAGARYIQMLITIGRTGYIYDHSPLSTWNAOG 967  
 QY 76 LRDLAAVEPVITPSDMEVKIITWGDYVACGDIISGLFVSARRREILLGADNFEQGW 135  
 DB 968 LRDLAAVEPVITPSDMEVKIITWGDYVACGDIISGLFVSARRREILLGADNFEQGW 1027  
 QY 136 RLIAPIYASQOTRGLLCIITSLTGRPKNOVEGVVSTATOSPLATCVNGVCTVEH 195  
 DB 1028 KLLAPITATOOTRGLGAIIVSLGRKXKNOAGOVULSSVOTFLGTSISGLVMTYH 1087  
 QY 196 GAGSXTLAGPKPIITOMTNNVDOLVGNQAPPGARSMTPTCGSSDLYLVTRHADVTPVR 255  
 DB 1088 GAGNTYLAGPKPIYOMTNTSAEGDLVGNPSPGTSLDPTCGADVTLVTRNADVTPVR 1147  
 QY 256 RRGDSRGLSLSPRPYSYLKSSGGELLCPGSHAVICIFPAACCTGVAAVDFIVESMET 315  
 DB 1148 RKDDRRGALLSPRLSTLKSSGGFVLCSRGHAGLFRFAACANGVAASIDFIVESLDV 1207  
 QY 316 TWRTSS 321  
 DB 1208 ATRTPS 1213  
 RESULT 9  
 ID DPO2 MOUSE STANDARD; PRT; 600 AA.  
 AC P33611;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)



```

FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409DBA CRC64;

Query Match 5.4%; Score 95.5; DB 1; Length 706;
Best Local Similarity 21.8%; Pred. No. 1.9;
Matches 67; Conservative 43; Mismatches 115; Indels 83; Gaps 18;

QY 48 MAFKLAALGTGY--YDHLTFLQDMAHAGRLDVAVEVPVIFSDMEVKIITWGA---100
DB 321 LGFRIIPAMDTWLYLGEYVT-----AIRLRIDIRPEVPKD-ECKKVKCAIGHH 371
QY 101 DTAACGP-IISGLPVARRGR-----ELLGPAINFEGQWRL-----LAPITAY 144
DB 372 EKVKCDKESVNSGAINICESHAQSTEDCTIAKIVGEALMSLDGFTYIACKGLVPLAE 431
QY 145 SQQTRGLGLITSLTRDKNOYGEVOVSTATQSFATCVNGVCTVTFHGAQKTLAG 204
DB 432 NYEIRSSACVDTEEGYH-----AAVAVKSSSDPLT-----NSLNG 470
QY 205 PKGPIITQMYTNVDOLVGMQAPPGARSMPTCTCGSSDLIYVTRHADVIPIRRGDSRGL 264
DB 471 KK-----SCHTGVND-TRAGMNIIMGL-----LYSEIKHCEPDFREGCAAPGYR 513
QY 265 LSPRPVYLKSSSGGP-LLC-PSGHA-----VGIFRAVCTRGVAKAVDFIPESME--T 315
DB 514 RNSLTLCLCTGCSAGCPRECEPNHERRYGTAFGLVEKGDVA---FVGHQYVEQNT 569
QY 316 TMRSSAW 323
DB 570 DGRNPDM 577

RESULT 11
VST2_HEVBU STANDARD; PRT; 660 AA.
AC P29326;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, last sequence update)
DT 01-FEB-1994 (Rel. 28, last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_Taxid=31767;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.W., Guerra M.E., Huang C.-C., Bradley D.W.,
RA Fry K.E., Reyes G.R.;
RA "Hepatitis E virus (HEV): molecular cloning and sequencing of the

```

```

RT full-length viral genome."
RL virology 185:120-131(1991).
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; W73218; AAA45736.1; -
DR PIR; C40778; YHMH2.
DR InterPro; IPR004261; SP2.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF03014; SP2; 1.
KM Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 660 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCD461C CRC64;

Query Match 5.3%; Score 93; DB 1; Length 660;
Best Local Similarity 19.0%; Pred. No. 2.8;
Matches 72; Conservative 47; Mismatches 119; Indels 140; Gaps 16;

QY 27 AQLIRACMLVRKKAAGHYVGMQAPFKLAALGTYYDHLTFLQDMAHAGRLDVAVEVP 86
DB 188 AATIRYRPLVFNAGVGAISISFPPQTTPTTSV-----DMNSTSTVDVILVQPG 239
QY 87 IFSDMEVKIITWGAADTAACDIIISGLPVARRGRILLGPAD--NFEQGWRLAPL-TA 143
DB 240 IASLVI-----PERRLHYNRQGRSVETSGVA 267
QY 144 YSQQTRGLL-----GCIIITSLTG-----161
DB 268 EREATSGVLMCIHSGIVNSYTNTPYTGAQLDLFALEFRNLPGNTNTRVRSYSTA 327
QY 162 --RDKNQYGEVOVSTATQSFIA---TCVNGV-----CMTVFN-----195
DB 328 RHLRLRGADGTAELTTTATRFKMDLYTSTNGVGEIGRGALTLFNLADTLGLPREL 387
QY 196 --GAG-----SKTLGAPRG-PITQMYTNVDOLVGMQAPPGARSMPTCTCGSSDLIYV-- 245
DB 388 ISSAGQLFYRRPVVANSANGFPTVLYTSVENA---QDDKGIATPHDIDLGESRVVIQDY 443
QY 246 --TRHADVIPIARRGDSRG-SILSRPVSYLK-----GSSGPIILCPSSHAYGIF 292
DB 444 DNQHQDPRPTSPAPSRPSVLRANDVLMLSLTAAEYDQSTYGSSTGQVTV--SDSVTLV 501
QY 293 RAAVCTRGVAKAVDFIPV 310
DB 502 NVATGAQAVARSLDWTKV 519

RESULT 12
VST2_HEVPA STANDARD; PRT; 660 AA.
AC P33426;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, last sequence update)
DT 01-FEB-1994 (Rel. 28, last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Pakistan) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_Taxid=33774;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,

```

RA Malik I.A., Iqbal M., Purcell R.H.;  
 RT "Characterization of a prototype strain of hepatitis E virus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSULATION OF THE GENOMIC RNA  
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M80581; AAA5727.1; -  
 DR InterPro: IPR004261; SP2.  
 DR InterPro: IPR008975; Viral\_cap\_coat.  
 DR Pfam: PF03014; SP2; 1.  
 KW Signal.  
 FT SIGNAL 1 22 BY SIMILARITY.  
 FT CHAIN 23 660 STRUCTURAL PROTEIN 2.  
 SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB46FD3 CRC64;  
 Query Match 5.3%; Score 93; DB 1; Length 660;  
 Best Local Similarity 19.0%; Pred. No. 2.8;  
 Matches 72; Conservative 49; Mismatches 117; Indels 140; Gaps 17;  
 QY 27 AAGIIRACMVRKAGGHVYOMAFMKLAALGTVYDHLPELOMAHAGRLDAVAVEPV 86  
 DB 188 ARATIRRLPLVNAVGVSAISISFTWPTTTPTSV-----DMSITSTVRIIVQPC 239  
 QY 87 IFSDMEVYITWGADTAACGDIISGLFVSARGREILLGPAD--NPEGQMRLL----- 138  
 DB 240 IASLVL-----PSERLHYNNQMRSVETSGVA 267  
 QY 139 -----AITAKSQT-RGLIGCI-----ITSITGDKNQ----- 166  
 DB 268 EEEATSGVLVLCIHGSPVNSTNTPTYGALGLDFALEFFRLNTPONTNTRVRSYSTA 327  
 QY 167 -----VEGEVQVSTATOSFLA-----TCVNGV-----CMTVFH----- 195  
 DB 328 RHRLRAGDGTALTTTAAIRFMDLVFTSTNGEIGRIGALITFLADTLTGLLPTL 387  
 QY 196 -GAG-----SKTAGRGK-PITOMTNNVDOLVGMQAPPGANSMPCTCGSSDLVYV-- 245  
 DB 388 ISSAGGQLFYSPRPVANSANGEPVTLVTSVENA-----QODKGIALPHDIDGESRVVIQDY 443  
 QY 246 -TRADVIVPVRRCDSRG-SLISRPVSYLK-----GSSGGPLLCFSGHVGIF 292  
 DB 444 DNGEHOORPTSPSPSPSVLNRANDVLMSTAAEDQSTYGSSTPVIYV--SDSVTLV 501  
 QY 293 RAAVCTRGVAKAVDFIV 310  
 DB 502 NVATGAQAVARSLDMTVY 519  
 RESULT 13  
 PTPV\_MOUSE STANDARD; PRT; 1705 AA.  
 ID PTPV\_MOUSE  
 AC P70289;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Receptor-type protein-tyrosine phosphatase v precursor (EC 3.1.3.48)  
 DE (Embryonic stem cell protein-tyrosine phosphatase) (ES cell  
 DE phosphatase).  
 GN PTPRV OR ESP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryonic stem cells;  
 RX MEDLINE=97109513; PubMed=8951793;  
 RA Lee K., Nichols U., Smith A.;  
 RT "Identification of a developmentally regulated protein tyrosine  
 RT phosphatase in embryonic stem cells that is a marker of  
 RT pluripotential epiblast and early mesoderm."  
 RL Mech. Dev. 59:153-164(1996).  
 RN [2]  
 RP ERRATUM.  
 RA Lee K., Nichols U., Smith A.;  
 RL Mech. Dev. 61:213-215(1996).  
 CC -1- FUNCTION: May play a role in the maintenance of pluripotency.  
 CC Down-regulated during differentiation.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DEVELOPMENTAL STAGES: Detectable in the epiblast of oocytes and  
 CC throughout early mouse embryo development. In adult, expression is  
 CC localized in gonadal germ cells.  
 CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
 CC -1- SIMILARITY: Contains 10 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U36488; AAC52868.1; -  
 DR HSSP: P18052; IYFO.  
 DR MGD: MGI:108027; Pcpv.  
 DR InterPro: IPR008957; FN\_III-like.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR000387; Tyr\_phosphatase.  
 DR InterPro: IPR000242; Tyr\_PP.  
 DR Pfam: PF00041; fn3; 7.  
 DR Pfam: PF00102; Y\_phosphatase; 1.  
 DR SMART: SMO0060; FN3; 8.  
 DR SMART: SMO0194; PTPC; 1.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE: PS00386; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE: PS0055; TYR\_PHOSPHATASE\_PTP; 2.  
 KW Hydroxylase; Transmembrane; Repeat; Signal; glycoprotein.  
 FT SIGNAL 1 18  
 FT CHAIN 19 1705  
 FT FT  
 FT DOMAIN 19 1077  
 FT TRANSMEM 1078 1100  
 FT DOMAIN 1101 1705  
 FT CYTOPLASMIC (POTENTIAL).  
 FT FIBRONECTIN TYPE-III 1.  
 FT FIBRONECTIN TYPE-III 2.  
 FT FIBRONECTIN TYPE-III 3.  
 FT FIBRONECTIN TYPE-III 4.  
 FT FIBRONECTIN TYPE-III 5.  
 FT FIBRONECTIN TYPE-III 6.  
 FT FIBRONECTIN TYPE-III 7.  
 FT FIBRONECTIN TYPE-III 8.  
 FT FIBRONECTIN TYPE-III 9.  
 FT FIBRONECTIN TYPE-III 10.  
 FT PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT PHOSPHOCYSTEINE INTERMEDIATE (BY  
 FT SIMILARITY).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 74 74  
 FT CARBOHYD 89 89  
 FT CARBOHYD 117 117  
 FT CARBOHYD 174 174  
 FT CARBOHYD 239 239  
 FT CARBOHYD 259 259

FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 663 663 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 970 970 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 982 982 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 SQ SEQUENCE 1705 AA; 186795 MM; 2783755F1587D5B CRC64;

Query Match 5.2%; Score 91.5; DB 1; Length 1705;  
 Best Local Similarity 25.7%; Pred. No. 12; Indels 99; Gaps 21;  
 Matches 83; Conservative 21; Mismatches 120;

65 LTPLOD--WNA---AGLRD---LAVAVEPIFSDMEVKIITWGDITACGDIISG-- 111  
 112 LTPLOD--WNA---AGLRD---LAVAVEPIFSDMEVKIITWGDITACGDIISG-- 162  
 113 LTPLOD--WNA---AGLRD---LAVAVEPIFSDMEVKIITWGDITACGDIISG-- 214  
 163 ---DKNOVEGEVQVSTATOSFLATCVNGVCTVFGAGSKTLGKPGPIITQMT--- 214  
 421 YSDAPAPLLNINISVSGATHTVFCGLVPGAHYV-----DIASMGDITQGLTGYTSP 473  
 215 -----NYDQDI--VQWQAPPGARSMTPTCTGSSSDIYLYTRADYIPVARRDSR- 261  
 474 LPPSLIIRNSPSDITLITMAPAP-----GMEGYKVTWHD-----GSQPS 516  
 262 -GSLISRP--VSYKSGSGGELLCPGSHAVGIF-----RAVCTRGVAKAVDPI 308  
 517 PGDIVDGPDISSITLTKSLVPGSC--YVTSANAMSGNLSDSQKHSCIRAPFNML 573  
 309 PVESMETTWTSSAMRHPOFGK 331  
 574 GFAHQPATLRAS--WCHPP--GGR 593

RESULT 14  
 VST2\_HEVME STANDARD; PRT; 659 AA.  
 ID VST2\_HEVME STANDARD; PRT; 659 AA.  
 AC Q03500;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor.  
 OS Hepatitis E virus (strain Mexico) (HEV).  
 OC Hepatitis E virus positive-strand viruses, no DNA stage;  
 OC Hepatitis E-like viruses.  
 OX NCBI\_Taxid=31768;  
 RN RP SEQUENCE FROM N.A.  
 RA MEDLINE=93079857; PubMed=144913;  
 RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,  
 RA Bradley D.W., Tam A.W., Reyes G.R.;  
 RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis  
 E virus (HEV)."  
 RT Virology 191:550-558 (1992).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 CC -2- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M74506; AAA45732.1; -  
 DR PIR; B44212; B44212; SP2.  
 DR InterPro; IPR004261; SP2.  
 DR InterPro; IPR008751; Viral\_cap\_coat.  
 DR Pfam; PF03014; SP2; 1.  
 KW Signal.  
 FT CHAIN 1 22 BY SIMILARITY.  
 FT SIGNAL 1 22 STRUCTURAL PROTEIN 2.  
 FT CHAIN 23 659  
 SQ SEQUENCE 659 AA; 70640 MM; C975E75EBD8B8E2C CRC64;

Query Match 5.1%; Score 90; DB 1; Length 659;  
 Best Local Similarity 17.6%; Pred. No. 5.2;  
 Matches 69; Conservative 54; Mismatches 124; Indels 146; Gaps 15;

27 AGLIRACMIVRKAGGHVYQMAFMKLAALGTYYVDTLPLODWAAGLRDLAVAEV 86  
 188 AATIRPPLVPAVAGVATISFTPTTPTSV-----DMNSITSDVRIIVQPG 239  
 87 IFSDMEVKIITWGDITACGDIISGLPVSARGREIILGPAD--NFGQGWRLI----- 138  
 240 IASELVI-----PSERLHYNQGRSVETSGVA 267  
 139 -----APTAVSQ-----OTRGLGCIITSLTGR----- 162  
 268 EBEATSGIWMCIHGSPNSTTTPYTGALGLDLFALEFRILTTNTNTRVRSYSTA 327  
 163 ---DKNOVEGEVQVSTATOSFLA---TCVNGV-----CMTVFGAGSKTL 202  
 338 RHSARGADGTALMTTATATRMKDLHFTGLNGVEVEGRGIALTLNLTADITLGLPYELI 387  
 203 AGPKG-----PTQMTYNVDQDVQWQAPPGARSMTPTCTGSSSDIYLYV--- 245  
 388 SSAGQLFYSPRPVVSANGPFPVKLYTSVENA---QDDKGVALPHDIDLDGRVVIDDYD 443  
 246 TRHADVIFVRRGRGSRG--SLSPRPVSYLK-----GSSGCPILCPGSHAVGIFR 293  
 444 NQHEQDRTPSPARSRPFSVLRANDVLMSTLAIEYDOSTGSGTGYVI--SDSVTLVN 501  
 294 AAVCTRGVAKAVDPI-----PVESMETTWT 319  
 502 VATGAQAVARSLDWSKXTLDGRPLPVEQVSKT 534

RESULT 15  
 POLG\_TBSEV STANDARD; PRT; 3414 AA.  
 ID POLG\_TBSEV STANDARD; PRT; 3414 AA.  
 AC P14336; Q08493;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Genome polyprotein (contains: Capsid protein C (Core protein); Matrix  
 protein (Envelope protein M); Major envelope protein E; Nonstructural  
 proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase  
 (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)  
 (NS5)).  
 OS Tick-borne encephalitis virus (Western subtype) (TBEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_Taxid=11088;  
 RN RP SEQUENCE FROM N.A., AND REVISIONS.  
 RA STRAIN=Neudoerfl;  
 RA MEDLINE=9603491; PubMed=7483260;  
 RA Wallner G., Mandl C.W., Kunz C., Heinz F.X.;  
 RT "The flavivirus 3'-noncoding region: extensive size heterogeneity  
 independent of evolutionary relationships among strains of tick-borne  
 encephalitis virus."  
 RT Virology 213:169-178 (1995).  
 RN [2]  
 RP SEQUENCE OF 1-779 FROM N.A.





```
FT STRAND 460 467
FT HELIX 468 470
FT TURN 475 476
FT STRAND 477 481
FT TURN 484 485
FT TURN 487 488
FT STRAND 492 496
FT HELIX 497 501
FT TURN 502 508
FT STRAND 507 508
FT TURN 510 511
FT STRAND 516 516
FT TURN 517 517
FT HELIX 518 521
FT STRAND 522 524
FT STRAND 526 528
FT TURN 529 530
FT TURN 531 531
FT STRAND 534 536
FT STRAND 541 547
FT HELIX 548 550
FT TURN 553 557
FT STRAND 558 559
FT TURN 560 562
FT STRAND 567 573
FT TURN 575 576
FT STRAND 577 577
FT TURN 582 583
FT STRAND 586 586
```

Query Match 5.0%; Score 88; DB 1; Length 3414;

Best Local Similarity 23.0%; Pred. No. 59;

Matches 59; Conservative 28; Mismatches 86; Indels 84; Gaps 12;

```
QY 64 HLPPLD-----WAHAGRLDVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVA 116
DB 1442 HLTLEKEEMKMAFWLIAGL-----AASAIHMSGIIGVMGLWLTLEMLRSSRSDLVFSG 1496
QY 117 RRGREILLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLGRDKNOVGEVOVST 176
DB 1497 QGGERGRDRPEYVDGV-YRIFSPGLFWGQ-----NOVG-----VGX 1532
QY 177 ATQSFATCVNGVWTFHAG---SKTLGPKGPITQWTVNVDOLV-----GM 223
DB 1533 GSKGVLT-----MMHVTGALSTIDDAVAGP-----YADVREDVVCYGAMSLDEKN 1581
QY 224 QA-----PCGARSMTPTCGSSDLVLYTRHADVI PVRRGDSRGSLSPPRVASYLKG 275
DB 1582 KGETVOYHAPPG-RAHEVHQCGGELIDT-----GRKLGAIPIDLVKG 1625
QY 276 SSGGPLLCPSGHAYGIF 292
DB 1626 TSGSPILNAOGVWVGLY 1642
```

Search completed: May 6, 2004, 09:31:48  
Job time : 9.97286 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 31.2384-Seconds

Title: US-10-650-585-4  
Perfect score: 1771  
Sequence: 1 MKKKLEHHHHHTSAGITK.....TIVFTSSAMWBPQFGKKKK 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Database : SPTREMBL\_25: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1551	87.6	3010	12	Q9J3H7	Q9J3H7 hepatitis c
2	1546	87.3	3010	12	Q68826	Q68826 hepatitis c
3	1546	87.3	3010	12	P90191	P90191 hepatitis c
4	1545	87.2	3010	12	Q9DTE6	Q9DTE6 hepatitis c
5	1543	87.1	3010	12	Q9DTE6	Q9DTE6 hepatitis c
6	1542	87.1	3010	12	Q9DTE6	Q9DTE6 hepatitis c
7	1541	87.0	3010	12	P88803	P88803 hepatitis c
8	1540	87.0	3010	12	Q9J3H5	Q9J3H5 hepatitis c
9	1540	87.0	3010	12	Q60735	Q60735 hepatitis c
10	1538	86.8	3010	12	Q9J3F9	Q9J3F9 hepatitis c
11	1535	86.7	361	12	C70815	C70815 hepatitis c
12	1535	86.7	3008	12	Q9J3F4	Q9J3F4 hepatitis c
13	1535	86.7	3008	12	Q9J3H3	Q9J3H3 hepatitis c
14	1534	86.6	3010	12	Q9J3H2	Q9J3H2 hepatitis c
15	1534	86.6	3010	12	Q9J3H0	Q9J3H0 hepatitis c
16	1533	86.6	3010	12	Q9J3Y3	Q9J3Y3 hepatitis c

17	1532	86.5	30.0	12	Q9Q1Y7	Q9Q1Y7	hepatitis c
18	1532	86.5	30.0	12	Q9Q1Y8	Q9Q1Y8	hepatitis c
19	1532	86.5	30.0	12	Q9Q1X6	Q9Q1X6	hepatitis c
20	1532	86.5	30.0	12	Q9Q1X5	Q9Q1X5	hepatitis c
21	1532	86.5	30.0	12	Q9Q3H6	Q9Q3H6	hepatitis c
22	1532	86.5	30.0	12	Q09796	Q09796	hepatitis c
23	1531	86.4	30.0	12	Q91AV0	Q91AV0	hepatitis c
24	1531	86.4	30.0	12	Q9J3H9	Q9J3H9	hepatitis c
25	1531	86.4	30.0	12	Q9WMX2	Q9WMX2	hepatitis c
26	1530	86.4	30.0	12	Q9J3H4	Q9J3H4	hepatitis c
27	1529	86.3	1186	12	Q81755	Q81755	hepatitis c
28	1529	86.3	2284	12	Q81817	Q81817	hepatitis c
29	1529	86.3	30.0	12	Q68788	Q68788	hepatitis c
30	1529	86.3	30.0	12	P89566	P89566	hepatitis c
31	1529	86.3	30.0	12	Q9DMD7	Q9DMD7	hepatitis c
32	1529	86.3	30.0	12	Q9DPE0	Q9DPE0	hepatitis c
33	1529	86.3	30.0	12	Q9DTE0	Q9DTE0	hepatitis c
34	1528	86.3	30.0	12	Q994U2	Q994U2	hepatitis c
35	1528	86.3	30.0	12	Q9DRE3	Q9DRE3	hepatitis c
36	1527	86.2	361	12	Q70818	Q70818	hepatitis c
37	1527	86.2	30.0	12	Q9QBP1	Q9QBP1	hepatitis c
38	1527	86.2	30.0	12	Q9J3G6	Q9J3G6	hepatitis c
39	1526	86.2	30.0	12	Q8Q2L8	Q8Q2L8	hepatitis c
40	1525	86.1	30.0	12	Q9Q3I1	Q9Q3I1	hepatitis c
41	1524	86.1	30.0	12	Q9Q1Y5	Q9Q1Y5	hepatitis c
42	1524	86.1	30.0	12	Q9DRE9	Q9DRE9	hepatitis c
43	1524	86.1	30.0	12	Q9Q1Y4	Q9Q1Y4	hepatitis c
44	1524	86.1	30.0	12	Q81541	Q81541	hepatitis c
45	1523	86.0	30.0	12	Q9J7G3	Q9J7G3	hepatitis c

## ALIGNMENTS

ID	09J3H7	PRELIMINARY;	PRT; 3010 AA.
AC	09J3H7;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DZ	Genome polypeptid.		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.		
OX	NCBI_TaxID=11103;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=KD15;		
RA	Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.		
RT	"Characteristics of hepatitis C viral genome associated with disease progression.";		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBPJ databases.		
CC	-1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS.		
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF		
CC	PROTEIN C AND RNA (BY SIMILARITY).		
DR	EMBL; AF207756; AA65946.1; -.		
DR	PIR; A61196; A61196.		
DR	PIR; P00246; P00246.		
DR	PIR; P00804; P00804.		
DR	PIR; PS0329; PS0329.		
DR	HSPF; P26663; IUXP.		
DR	GO; GO:0016021; C:Integral to membrane; IEA.		
DR	GO; GO:0019028; C:viral capsid; IEA.		
DR	GO; GO:0019031; C:viral envelope; IEA.		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0008026; F:ATP dependent helicase activity; IEA.		
DR	GO; GO:0005489; F:electron transporter activity; IEA.		
DR	GO; GO:0003723; F:RNA binding; IEA.		
DR	GO; GO:0003668; F:RNA-directed RNA polymerase activity; IEA.		
DR	GO; GO:0008336; F:serine-type peptidase activity; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		

DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro; IPR000345; Cys\_Ser\_lypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_core.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5a.  
 DR InterPro; IPR004109; peptidase C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 MW; D8653F7317FFA106 CRC64;

Query Match 87.6%; Score 1551; DB 12; Length 3010;  
 Best Local Similarity 95.4%; Pred. No. 4.6e-126;  
 Matches 290; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

DR 16 AGITKVFYFVAQGLIRACMLVRKAGHYVVAEMTALVTGTYVDHILTPIQDVAHAG 75  
 |||||  
 DB 904 AGIRMPYFVAQGLIRACMLVRKAGHYVVAEMTALVTGTYVDHILTPIQDVAHAG 963

DR 76 LRDLAVAVPPIFEDMEVKITITWADPAACGDIISGLPVARRGREITLGPANFEGQGM 135  
 |||||  
 DB 964 LRDLAVAVPPIFEDMEVKITITWADPAACGDIISGLPVARRGREITLGPANFEGQGM 1023

DR 136 RLAPITAYVSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCTVPH 195  
 |||||  
 DB 1024 RLAPITAYVSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCTVPH 1083

DR 196 GAGSKITLAGPKPTTQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLYVTRHADVPVR 255  
 |||||  
 DB 1084 GAGSKITLAGPKPTTQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLYVTRHADVPVR 1143

DR 256 RRGSRGSLSPRPVSYLKSGSGPILCPSGHAGIFRAAVCTRGVAKADVPVSEMET 315  
 |||||  
 DB 1144 RRGSRGSLSPRPVSYLKSGSGPILCPSGHAGIFRAAVCTRGVAKADVPVSEMET 1203

DR 316 TMRT 319  
 |||||  
 DB 1204 TMRT 1207

DR 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-U33;  
 RA Cho M.J.;  
 RT "Molecular cloning of Hepatitis C virus genome from a single Japanese  
 patient.";  
 RL Submitted (SEP-1991) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; D14484; BA03575.1; -;  
 DR PIR; A61196; A61196.  
 DR PIR; P0246; P0246.  
 DR PIR; P0804; P0804.  
 DR PIR; P80329; P80329.  
 DR HSRP; P26663; LUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_core.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5a.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR004109; peptidase C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR PRODOM; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM Hydrolyase; Nonstructural protein; Polyprotein;  
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327150 MW; 7270F47984554FAD CRC64;

Query Match 87.3%; Score 1546; DB 12; Length 3010;  
 Best Local Similarity 96.1%; Pred. No. 1.3e-125;

Matches 292; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITVPPFVRAQGLIRACMLVRKAGGHVYQMAFMKLAALGTIVYDHLTPLODMAHAG 75

DB 904 AGITVPPFVRAQGLIRACMLVRKAGGHVYQMAFMKLAALGTIVYDHLTPLODMAHAG 963

QY 76 LRDIAVAEPVIFSPMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 135

DB 964 LRDIAVAEPVIFSPMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 1023

QY 136 RLAPITVAYSOQTRGLLCIITSLTGRDKNVEGEVQVSTATOSFLATCVNGVCMVTFH 195

DB 1024 RLAPITVAYSOQTRGLLCIITSLTGRDKNVEGEVQVSTATOSFLATCVNGVCMVTFH 1083

QY 196 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPVR 255

DB 1084 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPVR 1143

QY 256 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMET 315

DB 1144 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMET 1203

QY 316 TMR 319

DB 1204 TMR 1207

RESULT 3

ID P90191 PRELIMINARY; PRT; 3010 AA.

AC P90191;

DT 01-MAY-1997 (TREMBlrel. 03, Created)

DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Genome polypeptide.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OC NCBI\_TaxID=11103;

OX [1]

RN SEQUENCE FROM N.A.

RC STRAIN=HCV-1b;

RA Enomoto N.; Asahina Y., Kurosaki M., Murakami T., Yamamoto C., Izumi N., Maruno F., Sato C.; "Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b.";

RT J. Clin. Invest. 96:224-230 (1995).

RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).

CC EMBL; D50482; BA05073.1; -.

DR PIR; A61196; A61196.

DR PIR; P00254; P00254.

DR PIR; P00804; P00804.

DR PIR; P0329; P0329.

DR PDB; 1DXV; 12-JAN-01.

DR GO; GO:0016021; C: integral to membrane; IEA.

DR GO; GO:0019028; C: viral capsid; IEA.

DR GO; GO:0019031; C: viral envelope; IEA.

DR GO; GO:0005524; F: ATP binding; IEA.

DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.

DR GO; GO:0003723; F: RNA binding; IEA.

DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F: serine-type peptidase activity; IEA.

DR GO; GO:0005198; F: structural molecule activity; IEA.

DR GO; GO:0016740; F: transferase activity; IEA.

DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P: transcription; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.

DR GO; GO:0019087; P: viral transformation; IEA.

DR Interpro; IPR009003; Cys Ser tryptsin.

DR Interpro; IPR001410; DEAD.

DR Interpro; IPR002522; HCV capsid.

DR Interpro; IPR002521; HCV capsid.

DR Interpro; IPR002519; HCV env.

DR Interpro; IPR002531; HCV NS1.

DR Interpro; IPR002518; HCV NS2.

DR Interpro; IPR000745; HCV NS4a.

DR Interpro; IPR001490; HCV NS4b.

DR Interpro; IPR002868; HCV NS4b.

DR Interpro; IPR002166; HCV NS5A.

DR Interpro; IPR001650; Helicase\_C.

DR Interpro; IPR004109; Peptidase\_C29.

DR Interpro; IPR007095; RNA pol ds ps.

DR Interpro; IPR007094; RNA pol ps vir.

DR Pfam; PF01543; HCV capsid; 1.

DR Pfam; PF01542; HCV core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5A; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR Pfam; PF00998; Viral RdRp; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; transferase; Transmembrane.

KW Polypeptide; RNA-directed RNA polymerase; transferase; Transmembrane.

FT CHAIN 1 191

FT CHAIN 192 383

FT CHAIN 384 809

FT CHAIN 810 1026

FT CHAIN 1027 1657

FT CHAIN 1658 1711

FT CHAIN 1712 1972

FT CHAIN 1973 2419

FT CHAIN 2420 3010

FT CHAIN NS5B.

FT CHAIN 3010

SC SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match 87.3%; Score 1546; DB 12; Length 3010;

Best Local Similarity 95.7%; Pred. NO. 1.3e-125;

Matches 291; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 16 AGITVPPFVRAQGLIRACMLVRKAGGHVYQMAFMKLAALGTIVYDHLTPLODMAHAG 75

DB 904 AGITVPPFVRAQGLIRACMLVRKAGGHVYQMAFMKLAALGTIVYDHLTPLODMAHAG 963

QY 76 LRDIAVAEPVIFSPMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 135

DB 964 LRDIAVAEPVIFSPMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGW 1023

QY 136 RLAPITVAYSOQTRGLLCIITSLTGRDKNVEGEVQVSTATOSFLATCVNGVCMVTFH 195

DB 1024 RLAPITVAYSOQTRGLLCIITSLTGRDKNVEGEVQVSTATOSFLATCVNGVCMVTFH 1083

QY 196 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPVR 255

DB 1084 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPVR 1143

QY 256 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMET 315

DB 1144 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMET 1203

QY 316 TMR 319

DB 1204 TMR 1207

RESULT 4  
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTE6; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV142;  
 RA Takahashi T., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.";  
 RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC LIPOPROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; AB049091; BAB18804.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; IUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0006118; F:electron transport; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR GO; GO:0003403; Cys Ser. trypsin.  
 DR InterPro; IPR00345; CytC\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5a.  
 DR InterPro; IPR001850; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVLr.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC\_C; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM Hydroxylase; Nonstructural protein; Polypeptide; Transmembrane.  
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 87.2%; Score 1545; DB 12; Length 3010;  
 Best Local Similarity 95.4%; Pred. No. 1.6e-125;  
 Matches 290; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 16 AGITKVPYFVPAQGLTRACMLVRAKAGHYVQWAFMKALATGYVYDHLTPIDNNAHAG 75  
 DB 904 AGITRVPYFVPAQGLTRACMLVRAKAGHYVQWAFMKALATGYVYDHLTPIDNNAHAG 963  
 QY 76 LRDIAVAVPEVTFSDMEVKIITWGDADPAAGDIISGLPVASARRGREILLGPADNFCQGV 135  
 DB 964 LRDIAVAVPEVTFSDMEVKIITWGDADPAAGDIISGLPVASARRGREILLGPADNFCQGV 1023  
 QY 136 RLAPITVYSCQITGLGCIITSLTGRDXQVVEGVQVSTATOSFIATCNGVGVFVH 195  
 DB 1024 RLAPITVYSCQITGLGCIITSLTGRDXQVVEGVQVSTATOSFIATCNGVGVFVH 1083  
 QY 196 GAGSKTLAGPKGPTTOMTYNDVODLVGMQAPPGARSMTPCTGSSDLYLTRHADVIPIVR 255  
 DB 1084 GAGSKTLAGPKGPTTOMTYNDVODLVGMQAPPGARSMTPCTGSSDLYLTRHADVIPIVR 1143  
 QY 256 RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRTGVAKANDFTVESMET 315  
 DB 1144 RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRTGVAKANDFTVESMET 1203  
 QY 316 TWRT 319  
 DB 1204 TWRS 1207

RESULT 5  
 Q9DTE4 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTE4; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV1150;  
 RA Takahashi T., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.";  
 RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC LIPOPROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; AB049093; BAB18806.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00804; P00804.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; IUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.

DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_tyrp\_sln.  
 DR InterPro: IPR000345; Cys\_Ser\_tyrp\_sln.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS2.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF01506; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00467; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SO SEQUENCE 3010 AA; 327324 MW; 3DE6CF249BD151C CRC64;

Query Match 87.1%; Score 1543; DB 12; Length 3010;  
 Best Local Similarity 95.4%; Pred. No. 2.3e-125;  
 Matches 290; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKYVYFRAQGLIRACMLVRAAGHYVMAFMALALGTYYVDHLTPIDQMAHAG 75  
 Db 904 AGITKYVYFRAQGLIRACMLVRAAGHYVMAFMALALGTYYVDHLTPIDQMAHAG 963  
 QY 76 LRDLAVAVEVITSDMEVKITITWAGDTAACGDIISGLPVARRGREITLGPADNFEQGM 135  
 Db 964 LRDLAVAVEVITSDMEVKITITWAGDTAACGDIISGLPVARRGREITLGPADNFEQGM 1023  
 QY 136 RLAPITAYSQQRGLIGCITITSLTGRDKQVGEVGVVSTAFQVATVAVCVMTVPH 195  
 Db 1024 RLAPITAYSQQRGLIGCITITSLTGRDKQVGEVGVVSTAFQVATVAVCVMTVPH 1083  
 QY 196 GAGSKTLAGEKPGITTYNTVNDLVGMQAPPGARSMPTCTGSSDLVLTTRADVIPIVR 255  
 Db 1084 GAGSKTLAGEKPGITTYNTVNDLVGMQAPPGARSMPTCTGSSDLVLTTRADVIPIVR 1143  
 QY 256 RRGDSRSGLSPRPVSLKXSSGGPILCPSGHVGITFRAVVCRTGAKAVDPIPVSMET 315  
 Db 1144 RRGDSRSGLSPRPVSLKXSSGGPILCPSGHVGITFRAVVCRTGAKAVDPIPVSMET 1203  
 QY 316 TMRIT 319  
 |||:

Db 1204 TMRS 1207  
 RESULT 6  
 QSDTD6 PRELIMINARY; PRT; 3010 AA.  
 ID QSDTD6  
 AC QSDTD6; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPRAIN=HCV221;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 Wshiro S.;  
 RT Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC EMBL: AB049101; BAB1814.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P26663; IUXP.  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0016787; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_tyrp\_sln.  
 DR InterPro: IPR000345; Cys\_Ser\_tyrp\_sln.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS2.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF01506; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.

```

DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00071; helicase_C; 1.
DR Pfam: PF00998; viral_RdRp; 1.
DR Pfam: PF186062; HCV_NSL; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA, 327108 MW, D8162D810EF78E84 CRC64;

Query Match 87.1%; Score 1542; DB 12; Length 3010;
Best Local Similarity 95.7%; Ped. No. 2,8e-125;
Matches 291; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYVRAQGLIRACMLVRKKAAGHYVMAMFKLAALGTIVYDHLTPLOPMALAG 75
DB 904 AVIHKVPYVRAQGLIRACMLVRKKAAGHYVMAMFKLAALGTIVYDHLTPLOPMALAG 963
QY 76 LMDLAVAEVPYFSDMEVKIITWGDPTAACGDIISGLFVYSARRGESILLGPDNFEQGGW 135
DB 964 LMDLVAEVPVFSMETKIIITWGDPTAACGDIISGLFVSARRGESILLGPDNFEQGGW 1023
QY 136 RLAAEITAVSQOTRGLGCIITSLTRDKNQVEGEVQVSTATOSFLATCNVGCWTFEH 195
DB 1024 RLAAEITAVSQOTRGLGCIITSLTRDKNQVEGEVQVSTATOSFLATCNVGCWTFEH 1083
QY 196 GAGSTIAGPKRPIQMTYTNVDGLVGNQAPPGARSMTPTCGSSDLYLVTNRADVLPVR 255
DB 1084 GAGSTIAGPKRPIQMTYTNVDGLVGNQAPPGARSMTPTCGSSDLYLVTNRADVLPVR 1143
QY 256 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHANGVIFPAAVCTRGAVAKVDFIVESMET 315
DB 1144 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHANGVIFPAAVCTRGAVAKVDFIVESMET 1203
QY 316 TWRT 319
DB 1204 TWRT 1207

RESULT 7
P8803 PRELIMINARY: PRT: 3010 AA.
AC P8803:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitis.
OX NCBI_TaxId=1103;
RN (1)
RP SEQUENCE FROM N.A.
RA Ectomoto N.;
RA STRAIN=HCV-1b;
RN Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA STRAIN=HCV-1b;
RX MEDLINE=95340824; PubMed=7542273;
RA Ectomoto C.; Sakuma T.; Asahina Y.; Kurosaki M.; Murakami T.;
RA Yamamoto C.; Izumi N.; Matsumoto F.; Sato C.;
RT "Comparison of full-length sequences of interferon-sensitive and
RT resistant hepatitis C virus 1b."
RL J. Clin. Invest. 96:224-230(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
EMBL: D50884; BAA09075.1; -.
PIR: A61196; A61196.

```

	DR	HSSP: P26663; UNS3.	C:integral to membrane; IEA.
	DR	GO: GO:0016021;	C:viral capsid; IEA.
	DR	GO: GO:0019028;	C:viral envelope; IEA.
	DR	GO: GO:0019031;	C:ATP binding; IEA.
	DR	GO: GO:0005543;	F:ATP dependent helicase activity; IEA.
	DR	GO: GO:0008026;	F:ATP dependent helicase activity; IEA.
	DR	GO: GO:0003723;	F:RNA binding; IEA.
	DR	GO: GO:0003968;	F:RNA-directed RNA polymerase activity; IEA.
	DR	GO: GO:0008265;	F:serine-type peptidase activity; IEA..
	DR	GO: GO:0005198;	F:structural molecule activity; IEA.
	DR	GO: GO:0016740;	F:transferase activity; IEA.
	DR	GO: GO:0006508;	P:proteolysis and peptidolysis; IEA.
	DR	GO: GO:0019079;	P:vital genome replication; IEA.
	DR	GO: GO:0019087;	P:vital transformation; IEA.
	DR	InterPro: IPRO09003;	Cys_Set_typepin.
	DR	InterPro: IPRO02410;	DEAD.
	DR	InterPro: IPRO02552;	HCV capsid.
	DR	InterPro: IPRO02521;	HCV core.
	DR	InterPro: IPRO02519;	HCV env.
	DR	InterPro: IPRO02531;	HCV NS1.
	DR	InterPro: IPRO02518;	HCV NS2.
	DR	InterPro: IPRO00745;	HCV NS4a.
	DR	InterPro: IPRO01490;	HCV NS4b.
	DR	InterPro: IPRO02868;	HCV NS5a.
	DR	InterPro: IPRO02166;	HCV RdRp.
	DR	InterPro: IPRO01650;	Helicase_C.
	DR	InterPro: IPRO04109;	Peptidase_C2.
	DR	InterPro: IPRO07095;	RNA_pol_DS_PS.
	DR	InterPro: IPRO07094;	RNA_pol_PSVLR.
	DR	Pfam: PF01543;	HCV capsid; 1.
	DR	Pfam: PF01539;	HCV env; 1.
	DR	Pfam: PF01560;	HCV NS1; 1.
	DR	Pfam: PF01538;	HCV NS2; 1.
	DR	Pfam: PR02907;	HCV NS3; 1.
	DR	Pfam: PF01006;	HCV NS4a; 1.
	DR	Pfam: PF01001;	HCV NS4b; 1.
	DR	Pfam: PF01506;	HCV NS5a; 1.
	DR	Pfam: PF00271;	helicase_C; 1.
	DR	Pfam: PF00998;	viral RdRp; 1.
	DR	Prodont; PD186062;	HCV NS1; 1.
	DR	SMART; SMO0487;	DExdcj; 1.
	KM	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-direccted RNA polymerase; Transferase; Transmembrane.	
	FT	CHAIN	1 181 CORE PROTEIN.
	FT	CHAIN	192 383 E1.
	FT	CHAIN	384 809 E2.
	FT	CHAIN	810 1026 NS2.
	FT	CHAIN	1027 1657 NS3.
	FT	CHAIN	1658 1711 NS4a.
	FT	CHAIN	1712 1972 NS4B.
	FT	CHAIN	1973 2419 NS5A.
	FT	CHAIN	2420 3010 NS5B.
	SQ	SEQUENCE	3010 AA; 327332 KM; 5F81505783FEFFB8 CRC64;
	Query Match	87.0%; Score 1541; DB 12; Length 3010;	
	Best Local Similarity	95.1%; Pred. No.3.se-125;	
	Matches 289; Conservative 10; Mismatches 5; Indels 0; Gaps 0;		
OY	16 AGITKPYVRAROGIIRACMCYRKAAGHYGMAFMKLAAALGTGYVPHLPIDWMHAG	75	
DB	904 AGIRRPPIFRQGILRCMLWRKVAGHYQAFAFKLALIGTYVINHLPLDQMARTG	963	
OY	76 LRDLVAVEPVLFSDMEVKIIITWGADTTAACGDIIISGLPVARSGRSEILLGPADNEEGGW	135	
DB	964 LRDLVAVEVPVFSDMEETKIITWGADTTPAACGDIIISGLPVASARGREIILLGPADSPEGGW	1023	
OY	136 RLIAPIITYSOTCTGGICITITSITSGDDKNUGGEVGWYSTQTGSFLATCNVGVCMTFEH	195	
DB	1024 RLIAPIITYSQCTRLKGCIITSITGDKDNUGGEVGWYSTAQSLATCNVGCMTVIYH	1063	
OY	196 GAGSKTLGAPRKPTITWTYNVDOLVGMQAPGARSMETCCTGSSLDYLVTRHADVIVPR	255	



```

Db      1084 GAGSKTLAGPGPIOMTYNVDOLVGMQAPPGARSLLPTCCGSSDLVLTVRHADVIPVR 1143
QY      256 RRGDSRGSLLSPRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDIPVESMET 315
Db      1144 RRGDSRGSLLSPRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDIPVESMET 1203
QY      316 TMRT 319
Db      1204 TMS 1207

RESULT 8
ID Q9J3H5 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DB Genome polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD17;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; AF207758; AAF65948.1; -.
DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; P00254; P00254.
DR PIR; P00329; P00329.
DR HSBP; P27958; HEB1.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F: electron transporter activity; IEA.
DR GO; GO:0016787; F: hydrolase activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR GO; GO:0016740; F: transferase activity; IEA.
DR GO; GO:0006118; F: electron transport; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P: transcription; IEA.
DR GO; GO:0019079; P: viral genome replication; IEA.
DR GO; GO:0019087; P: viral transformation; IEA.
DR InterPro: IPR009003; Cys_Ser_tyrpstin.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV_NS4.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR0041650; Helicase_C.
DR InterPro: IPR004109; Peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.

```

```

DR InterPro: IPR007095; RNA_pol_PsVlr.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326501 MW; 9FE83D1B93B7AA48 CRC64;

Query Match 87.0%; Score 1540; DB 12; Length 3010;
Best Local Similarity 95.4%; Pred. No. 4.2e-125;
Matches 290; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYVRAGGLIRACMLYRKAGHYVOMAFKLAALTGTYYDHLTLPQMAHAG 75
Db 904 AGITRVPYVRAGGLIRACMLYRKAGHYVOMAFKLAALTGTYYDHLTLPQMAHAG 963
QY 76 LRDLAAVAPVTFSDMEVITITWGADTAACGDIISGLPSARGREITLGPADNFGGGM 135
Db 964 LRDLAAVAPVTFSDMETIITWGADTAACGDIISGLPSARGREITLGPADNFGGGM 1023
QY 136 RLIAPIYVSSQTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCVNGVCMYVR 195
Db 1024 RLIAPIYVSSQTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCVNGVCMYVR 1083
QY 196 GAGSKTLAPKPKPIOMTYNVDOLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPR 255
Db 1084 GAGSKTLAPKPKPIOMTYNVDOLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPR 1143
QY 256 RRGDSRGSLLSPRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDIPVESMET 315
Db 1144 RRGDSRGSLLSPRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDIPVESMET 1203
QY 316 TMRT 319
Db 1204 TMS 1207

RESULT 9
ID Q807B3 PRELIMINARY; PRT; 3010 AA.
AC Q807B3;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DB Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M1E;
RA MEDLINE=22047193; PubMed=12051758;
RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
RA Nio Y., Hosaka Y., Miyahara Y., Shimotohno K.;
RT "Subgenomic replicon derived from a cell line infected with the
RT hepatitis C virus."
RL Biochem. Biophys. Res. Commun. 293:993-999 (2002).
DR EMBL; AB080299; BAC54896.1; -.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.

```



DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; viral\_RBP; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; Transference; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C9B93E60C7 CRC64;

Query Match 86.8%; Score 1538; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 6.4e-125;  
 Matches 288; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 75  
 DB 904 AGITRMPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 963  
 QY 76 LRDLAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 135  
 DB 964 LRDLAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 1023  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCMTVFH 195  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCMTVFH 1083  
 QY 196 GAGSKTLAPKGPITQMTYTNVDQLVGQAPPGARSMPTCTCGSSDLVYTRHADVPVR 255  
 DB 1084 GAGSKTLAPKGPITQMTYTNVDQLVGQAPPGARSMPTCTCGSSDLVYTRHADVPVR 1143  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMET 315  
 DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 11  
 ID 070815 PRELIMINARY; PRT; 361 AA.  
 AC 070815;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polypeptide (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98321154; PubMed=9656998;  
 RA Yamada K., Mori A., Seki M., Kimura J., Yuasa S., Matsuura Y.,  
 RA Miyamura T.,  
 RT "Critical point mutations for hepatitis C virus NS3 proteinase.",  
 RL Virology 246:104-112(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,  
 RA Miyamura T.,  
 RT "Enzymatic characterization of purified NS3 serine proteinase of  
 RT hepatitis C virus expressed in Escherichia coli.",  
 RL FEBS Lett. 378:37-42(1998).  
 DR EMBL; AB013620; BAA28498.1; -.  
 DR HSSP; P27958; 1HEI.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_cypsin.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.  
 FT NON\_TER 1  
 FT TER 361  
 SQ SEQUENCE 361 AA; 38336 MW; 87DC310C76F4BC3 CRC64;  
 Query Match 86.7%; Score 1535; DB 12; Length 361;  
 Best Local Similarity 94.7%; Pred. No. 7.1e-126;  
 Matches 288; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 75  
 DB 5 AGITRMPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 64  
 QY 76 LRDLAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 135  
 DB 65 LRDLAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 124  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCMTVFH 195  
 DB 125 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCMTVFH 184  
 QY 196 GAGSKTLAPKGPITQMTYTNVDQLVGQAPPGARSMPTCTCGSSDLVYTRHADVPVR 255  
 DB 185 GAGSKTLAPKGPITQMTYTNVDQLVGQAPPGARSMPTCTCGSSDLVYTRHADVPVR 244  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMET 315  
 DB 245 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMET 304  
 QY 316 TMRT 319  
 DB 305 TMRS 308

RESULT 12  
 ID 0903F4 PRELIMINARY; PRT; 3008 AA.  
 AC 0903F4;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 GN MD34.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD34;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.,  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 RT progression.",  
 RL Submitted (Nov-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN M AND RNA (BY SIMILARITY).  
 DR EMBL; AF208024; AAF61205.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00245; P00245.  
 DR PIR; P80329; P80329.  
 DR HSSP; P26663; 1UXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005488; F:electron transporter activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro: IPR000345; CysC\_heme\_BS.  
 DR InterPro: IPR004110; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR Pfam: PD186062; HCV\_NS1; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEAD; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolyase; Nonstructural protein; Transmembrane;  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane;  
 SQ SEQUENCE 3008 AA; 326834 MM; 99AB09E1C3109F4 CAC64;

Query Match 86.7%; Score 1535; DB 12; Length 3008;  
 Best Local Similarity 95.4%; Pred. No. 1,2e-124;  
 Matches 290; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

CY 16 AGTKKVPYFVPAQGLIYACMLVRAAGHYVQVAFMTALITGYVYDHLTPLDMAHAG 75  
 DB 902 ASIKKVPYFVPAQGLIYACMLVRAAGHYVQVAFMTALITGYVYDHLTPLDMAHAG 961  
 CY 76 LRDLAVAVEPYFSDMEVKIITWGDATTAACDIIISGLPVASARGREILLPPADNFEQGM 135  
 DB 962 LRDLAVAVEPYFSDMEVKIITWGDATTAACDIIISGLPVASARGREILLPPADNFEQGM 1021  
 CY 136 RLAPITAYSOQTGLIGCITITSTGRDKQVEGEVQVSTATQSFATCVNVCMTVFH 195  
 DB 1022 RLAPITAYSOQTGLIGCITITSTGRDKQVEGEVQVSTATQSFATCVNVCMTVFH 1081  
 CY 196 GAGSKTLAAGKGPFTQMTYVNDQVLVGMQAPPGARSWTPTCCSSSDLYLTVRAADYIPVR 255  
 DB 1082 GAGSKTLAAGKGPFTQMTYVNDQVLVGMQAPPGARSWTPTCCSSSDLYLTVRAADYIPVR 1141  
 CY 256 RRGDSRGLSPRPVSYLKSGSGPPLICPSGHAAGIFRAAVCTRGVAKAVDFPVSMET 315  
 DB 1142 RRGDSRGLSPRPVSYLKSGSGPPLICPSGHAAGIFRAAVCTRGVAKAVDFPVSMET 1201  
 CY 316 TMR 319  
 DB 1202 TMR 1205

ID 09J3H3 PRELIMINARY; PRT: 3010 AA.  
 AC 09J3H3  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 DE Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MD19;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RT Submitted (Nov-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL AF207760; AAF5950.1; .  
 DR PIR A61196; A61196.  
 DR PIR PS0329; PS0329.  
 DR HSSP; P26663; IUXP.  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; F:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro: IPR000345; CysC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEAD; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.

SQ SEQUENCE 3010 AA; 327234 MW; 44C34677649C8BD CRC64;  
 Query Match 86.7%; Score 1535; DB 12; Length 3010;  
 Best Local Similarity 94.1%; Pred. No. 1.2e-124;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFRAQGLIRACMLVRKAAGHYVQNAFMLALALGTYYVDHLTPLODVAHAG 75  
 DB 904 AGITRVYFRAQGLIRACMLVRKAAGHYVQNAFMLALALGTYYVDHLTPLODVAHAG 963  
 QY 76 LRDLAVAVEPVFSDMEVKIITWGADTAACDIIISGLPVASARGREIILGPADNFEQGM 135  
 DB 964 LRDLAVAVEPVFSDMEVKIITWGADTAACDIIISGLPVASARGREIILGPADNFEQGM 1023  
 QY 136 RLAPITAYSCQTRGLIGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMTVFH 195  
 DB 1024 RLAPITAYSCQTRGLIGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMTVFH 1083  
 QY 196 GAGSKTLAEPKPIITOMYTNVDODLVGMQAPPGARSMTPTCCSSDLYLTVTRADVIPIVR 255  
 DB 1084 GAGSKTLAEPKPIITOMYTNVDODLVGMQAPPGARSMTPTCCSSDLYLTVTRADVIPIVR 1143  
 QY 256 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 14  
 Q9J3H2 PRELIMINARY; PRT; 3010 AA.  
 ID Q9J3H2  
 AC Q9J3H2  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CC NCBI\_Taxid=1103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MO20;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NCV-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL, AF207761; AAE5951.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; INS3.  
 DR GO; GO:0016031; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:000489; F: electron transporter activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003668; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.  
 DR GO; GO:0006118; P: electron transport; IEA.  
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro; IPR000345; CysC\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002668; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_D8\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326763 MW; 1A4BBE4BE5144DD0 CRC64;

Query Match 86.6%; Score 1534; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 1.4e-124;  
 Matches 288; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFRAQGLIRACMLVRKAAGHYVQNAFMLALALGTYYVDHLTPLODVAHAG 75  
 DB 904 AGITRVYFRAQGLIRACMLVRKAAGHYVQNAFMLALALGTYYVDHLTPLODVAHAG 963  
 QY 76 LRDLAVAVEPVFSDMEVKIITWGADTAACDIIISGLPVASARGREIILGPADNFEQGM 135  
 DB 964 LRDLAVAVEPVFSDMEVKIITWGADTAACDIIISGLPVASARGREIILGPADNFEQGM 1023  
 QY 136 RLAPITAYSCQTRGLIGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMTVFH 195  
 DB 1024 RLAPITAYSCQTRGLIGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMTVFH 1083  
 QY 196 GAGSKTLAEPKPIITOMYTNVDODLVGMQAPPGARSMTPTCCSSDLYLTVTRADVIPIVR 255  
 DB 1084 GAGSKTLAEPKPIITOMYTNVDODLVGMQAPPGARSMTPTCCSSDLYLTVTRADVIPIVR 1143  
 QY 256 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 15  
 Q9J3I0 PRELIMINARY; PRT; 3010 AA.  
 ID Q9J3I0  
 AC Q9J3I0  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Genome polyprotein.

OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD12;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC EMBL; AF207753; AAF65943.1; -  
 CC PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; IUXP.  
 DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F: electron transporter activity; IEA.  
 DR GO; GO:0016787; F: hydrolase activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.  
 DR GO; GO:0006118; F: electron transport; IEA.  
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.  
 DR InterPro; IPR003003; Cys Ser lyso. Cys.  
 DR InterPro; IPR00345; Cys Ser lyso. Cys.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD16062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 KM ATP-binding, Coat protein, Envelope protein, Glycoprotein, Helicase;  
 KM Hydrolase, Nonstructural protein, Polypeptide, Transmembrane,  
 KM RNA-directed RNA polymerase, Transferase, Transmembrane.  
 SQ SEQUENCE 3010 AA; 32692 MW; 074098D305AFLA9 CRC64;

Query Match 86.6%; Score 1534; DB 12; Length 3010;

Best Local Similarity 95.1%; Pred. No. 1.4e-124;  
 Matches 289; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
 QY 16 AGITKVPYFVAQGLIRACMLVREKAGGHVYQVAFMTLALGTGYVDLTPIDMAHAG 75  
 DB 904 AGITRVPYFVAQGLIRACMLVREKAGGHVYQVAFMTLALGTGYVDLTPIDMAHAG 963  
 QY 76 LRDIAVAVEPIFDMEKXITTWGADTPACDIIISGIPVBARREIILGPADDFEGQW 135  
 DB 964 LRDIAVAVEPIFDMEKXITTWGADTPACDIIISGIPVBARREIILGPADDFEGQW 1023  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQVSTATOSFLATCVNGVMTVFH 195  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQVSTATOTFLATCVNGVMTVFH 1083  
 QY 196 GAGSKTLAEPKPIITOMTNTDVLVQAPRPARSMTPCTCGSSDIYVTRHADVTPVR 255  
 DB 1084 GAGSKTLAEPKPIITOMTNTDVLVQAPRPARSMTPCTCGSSDIYVTRHADVTPVR 1143  
 QY 256 RRGDSRGSILSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTGKAVKADFIPEESMET 315  
 DB 1144 RRGDSRGSILSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTGKAVKADFIPEESMET 1203  
 QY 316 TWRT 319  
 DB 1204 TWRS 1207

Search completed: May 6, 2004, 09:35:44  
 Job time : 34.2384 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13.109 Seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771  
Sequence: 1 MKKKKLEHHHHTSAGITK.....TTMKTSAMRHPQGGKKKK 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A\_COM3.pdp:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COM3.pdp:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COM3.pdp:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COM3.pdp:\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUTS\_COM3.pdp:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pdp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1531	86.4	2201	4 US-09-539-601-6	Sequence 6, Appl
2	1531	86.4	2201	4 US-09-539-601-15	Sequence 15, Appl
3	1531	86.4	3010	4 US-09-539-601-3	Sequence 3, Appl
4	1531	86.4	3010	4 US-09-539-601-21	Sequence 21, Appl
5	1531	86.4	3010	4 US-09-539-601-27	Sequence 27, Appl
6	1528	86.3	1692	3 US-09-263-933-4	Sequence 4, Appl
7	1528	86.3	1692	4 US-09-919-901-4	Sequence 4, Appl
8	1528	86.3	2307	3 US-09-263-933-2	Sequence 2, Appl
9	1528	86.3	2307	3 US-09-919-901-2	Sequence 2, Appl
10	1525	86.1	1692	3 US-09-263-933-11	Sequence 11, Appl
11	1525	86.1	1692	4 US-09-919-901-11	Sequence 11, Appl
12	1525	86.1	2307	3 US-09-263-933-9	Sequence 9, Appl
13	1525	86.1	2307	4 US-09-919-901-9	Sequence 9, Appl
14	1524	86.1	3010	4 US-09-539-601-33	Sequence 33, Appl
15	1516	85.6	1692	3 US-09-263-933-18	Sequence 18, Appl
16	1516	85.6	1692	4 US-09-919-901-18	Sequence 18, Appl
17	1516	85.6	2307	3 US-09-263-933-16	Sequence 16, Appl
18	1516	85.6	2307	4 US-09-919-901-16	Sequence 16, Appl
19	1505	85.0	3010	3 US-09-014-416-3	Sequence 3, Appl
20	1479	83.5	2013	1 US-08-324-977-12	Sequence 12, Appl
21	1479	83.5	2013	2 US-08-384-616-12	Sequence 12, Appl
22	1479	83.5	2013	2 US-08-904-686A-12	Sequence 12, Appl
23	1479	83.5	2013	3 US-09-315-850-12	Sequence 12, Appl
24	1479	83.5	2201	4 US-08-952-981A-2	Sequence 2, Appl
25	1479	83.5	2620	1 US-08-324-977-32	Sequence 32, Appl
26	1479	83.5	2620	2 US-08-384-616-32	Sequence 32, Appl
27	1479	83.5	2620	2 US-08-904-686A-32	Sequence 32, Appl

28	1479	83.5	2620	3 US-09-315-850-32	Sequence 32, Appl
29	1479	83.5	2621	1 US-08-324-977-36	Sequence 36, Appl
30	1479	83.5	2621	2 US-08-384-616-36	Sequence 36, Appl
31	1479	83.5	2621	2 US-08-904-686A-36	Sequence 36, Appl
32	1479	83.5	2621	2 US-09-315-850-36	Sequence 36, Appl
33	1479	83.5	3010	1 US-08-324-977-2	Sequence 2, Appl
34	1479	83.5	3010	1 US-08-324-977-14	Sequence 14, Appl
35	1479	83.5	3010	2 US-08-384-616-2	Sequence 2, Appl
36	1479	83.5	3010	2 US-08-384-616-14	Sequence 14, Appl
37	1479	83.5	3010	2 US-08-904-686A-2	Sequence 2, Appl
38	1479	83.5	3010	2 US-08-904-686A-14	Sequence 14, Appl
39	1479	83.5	3010	3 US-09-315-850-2	Sequence 2, Appl
40	1479	83.5	3010	3 US-09-315-850-14	Sequence 14, Appl
41	1406	79.4	3012	3 US-08-811-566-2	Sequence 2, Appl
42	1406	79.4	3012	4 US-09-034-756-2	Sequence 2, Appl
43	1405	79.3	2894	4 US-08-466-975A-23	Sequence 23, Appl
44	1405	79.3	2894	2 US-08-391-671A-23	Sequence 23, Appl
45	1405	79.3	2894	3 US-08-467-902A-23	Sequence 23, Appl

# ALIGNMENTS

RESULT 1  
US-09-539-601-6  
; Sequence 6, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Barten Schlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; EARLIER FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-6

Query Match 86.4%; Score 1531; DB 4; Length 2201;  
Best Local Similarity 94.7%; Pred. No. 2e-133; 7; Indels 0; Gaps 0;  
Matches 288; Conservative 9; Mismatches 7

QY	16	AGTTKVPYVRAQGLTRACMLVYKKAAGHYVOMAFKLAALGTGYVYDHLTPLQDMANAG	75
DB	95	AGTTKVPYVRAHGLTRACMLVYKKAAGHYVOMAFKLAALGTGYVYDHLTPLQDMANAG	154
QY	76	LRDLAAVPEVPIFSDVEVKITITMGADTACGDIISGLPVASARGREIILGPADNFEQGM	135
DB	155	LRDLAAVPEVPIFSDVEVKITITMGADTACGDIISGLPVASARGREIILGPADNFEQGM	214
QY	136	RLIAPITVAYSGQTRGLGCIITSLGRDNQYEGEVQVSTATQSFPLATCVGVCWTYVH	195
DB	215	RLIAPITVAYSGQTRGLGCIITSLGRDNQYEGEVQVSTATQSFPLATCVGVCWTYVH	274
QY	196	GGGSKTLAEPKPIITMTNNVDOLVGMQAPPGARSMPCTGSSSLVYVTHADVIPIR	255
DB	275	GGGSKTLAEPKPIITMTNNVDOLVGMQAPPGARSMPCTGSSSLVYVTHADVIPIR	334
QY	256	RRDSSGSLSPRPVSYLKGSSGGPLLCPSGAVGIFRAAVCTRGVAKAVDEIPVESMET	315
DB	335	RRDSSGSLSPRPVSYLKGSSGGPLLCPSGAVGIFRAAVCTRGVAKAVDEIPVESMET	394
QY	316	TRMT 319	
DB	395	TRMS 398	

RESULT 2



US-09-539-601-15  
; Sequence 15, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlagel, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-15

Query Match 86.4%; Score 1531; DB 4; Length 2201;  
Best Local Similarity 94.7%; Pred. No. 2e-143;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75  
DB 95 AGITKVPYFVRAHGLIRACMLVRKVAAGHYVQMALMLKLAALTGYVYDHLTPLODMAHAG 154  
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135  
DB 155 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 214  
QY 136 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVFH 195  
DB 215 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVFH 274  
QY 196 GAGSKTLGPKGPITQMTNTVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 255  
DB 275 GAGSKTLGPKGPITQMTNTVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 334  
QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 315  
DB 335 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 394  
QY 316 TMRT 319  
DB 395 TMS 396

RESULT 3  
US-09-539-601-3  
; Sequence 3, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlagel, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-3

Query Match 86.4%; Score 1531; DB 4; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 3.2e-143;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75

DB 904 AGITKVPYFVRAHGLIRACMLVRKVAAGHYVQMALMLKLAALTGYVYDHLTPLODMAHAG 963  
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135  
DB 964 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 1023  
QY 136 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVFH 195  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVFH 1083  
QY 196 GAGSKTLGPKGPITQMTNTVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 255  
DB 1084 GAGSKTLGPKGPITQMTNTVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 1143  
QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 315  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 1203  
QY 316 TMRT 319  
DB 1204 TMS 1207

RESULT 4  
US-09-539-601-21  
; Sequence 21, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlagel, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-21

Query Match 86.4%; Score 1531; DB 4; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 3.2e-143;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75  
DB 904 AGITKVPYFVRAHGLIRACMLVRKVAAGHYVQMALMLKLAALTGYVYDHLTPLODMAHAG 963  
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135  
DB 964 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 1023  
QY 136 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVFH 195  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVFH 1083  
QY 196 GAGSKTLGPKGPITQMTNTVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 255  
DB 1084 GAGSKTLGPKGPITQMTNTVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 1143  
QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 315  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 1203  
QY 316 TMRT 319  
DB 1204 TMS 1207

RESULT 5  
 US-09-539-601-27  
 ; Sequence 27, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartschlagel, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 27  
 ; LENGTH: 3010  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-27

Query Match 86.4%; Score 1531; DB 4; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 3.2e-143;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 75  
 |||||  
 DB 904 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 963  
 QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 135  
 |||||  
 DB 964 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 1023  
 QY 136 RLAPITAYSOOTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTVFH 195  
 |||||  
 DB 1024 RLAPITAYSOOTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTVFH 1083  
 QY 196 GAGSKTLGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 255  
 |||||  
 DB 1084 GAGSKTLGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 1143  
 QY 256 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHANGIRAVCTRGVAKADPIPVESMET 315  
 |||||  
 DB 1144 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHANGIRAVCTRGVAKADPIPVESMET 1203  
 QY 316 TMTT 319  
 |||||  
 DB 1204 TMTT 1207

RESULT 6  
 US-09-263-933-4  
 ; Sequence 4, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patrick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/263,933  
 ; CURRENT FILING DATE: 1999-03-08  
 ; EARLIER APPLICATION NUMBER: 09/129,611  
 ; EARLIER FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 US-09-263-933-4

Query Match 86.3%; Score 1528; DB 3; Length 1692;

Best Local Similarity 94.1%; Pred. No. 2.7e-143;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 75  
 |||||  
 DB 183 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 242  
 QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 135  
 |||||  
 DB 243 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
 QY 136 RLAPITAYSOOTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTVFH 195  
 |||||  
 DB 303 RLAPITAYSOOTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTVFH 362  
 QY 196 GAGSKTLGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 255  
 |||||  
 DB 363 GAGSKTLGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 422  
 QY 256 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHANGIRAVCTRGVAKADPIPVESMET 315  
 |||||  
 DB 423 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHANGIRAVCTRGVAKADPIPVESMET 482  
 QY 316 TMTT 319  
 |||||  
 DB 483 TMTT 486

RESULT 7  
 US-09-919-901-4  
 ; Sequence 4, Application US/09919901  
 ; Patent No. 6599738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patrick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 US-09-919-901-4

Query Match 86.3%; Score 1528; DB 4; Length 1692;  
 Best Local Similarity 94.1%; Pred. No. 2.7e-143;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 75  
 |||||  
 DB 183 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 242  
 QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 135  
 |||||  
 DB 243 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
 QY 136 RLAPITAYSOOTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTVFH 195  
 |||||  
 DB 303 RLAPITAYSOOTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTVFH 362  
 QY 196 GAGSKTLGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 255  
 |||||

Db 363 GAGSKTLAAGPKGPIITOMYTNVDODLVGMQAPPGARSITPCTCGSSDLYLVTRHADVI PVR 422  
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIGFRAAVCTRGVAKAVDFIPVESMET 315  
Db 423 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIGFRAAVCTRGVAKAVDFIPVESMET 482  
QY 316 TMRT 319  
Db 483 TMRS 486

RESULT 8  
US-09-263-933-2  
Sequence 2, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PR  
ORGANISM: Artificial Sequence  
US-09-263-933-2

Query Match 86.3%; Score 1528; DB 3; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 4.3e-143;  
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;  
QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQAFMKLAALTGTYYVDHLTPLODMAHAG 75  
Db 275 AGITRVYFVRAOGLIRACMLVRKAGHYVQAFMKLAALTGTYYVDHLTPLODMAHAG 334  
QY 76 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
Db 335 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 394  
QY 136 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 195  
Db 395 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 454  
QY 196 GAGSKTLAAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVI PVR 255  
Db 455 GAGSKTLAAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVI PVR 514  
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIGFRAAVCTRGVAKAVDFIPVESMET 315  
Db 515 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIGFRAAVCTRGVAKAVDFIPVESMET 574  
QY 316 TMRT 319  
Db 575 TMRS 578

RESULT 9  
US-09-919-901-2  
Sequence 2, Application US/09919901  
Patent No. 6599738  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692  
TYPE: PR  
ORGANISM: Artificial Sequence  
US-09-919-901-2

FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PR  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-2

Query Match 86.3%; Score 1528; DB 4; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 4.3e-143;  
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;  
QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQAFMKLAALTGTYYVDHLTPLODMAHAG 75  
Db 275 AGITRVYFVRAOGLIRACMLVRKAGHYVQAFMKLAALTGTYYVDHLTPLODMAHAG 334  
QY 76 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
Db 335 LRDLAAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 394  
QY 136 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 195  
Db 395 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 454  
QY 196 GAGSKTLAAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVI PVR 255  
Db 455 GAGSKTLAAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVI PVR 514  
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIGFRAAVCTRGVAKAVDFIPVESMET 315  
Db 515 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIGFRAAVCTRGVAKAVDFIPVESMET 574  
QY 316 TMRT 319  
Db 575 TMRS 578

RESULT 10  
US-09-263-933-11  
Sequence 11, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692  
TYPE: PR  
ORGANISM: Artificial Sequence  
US-09-263-933-11

Query Match 86.1%; Score 1525; DB 3; Length 1692;  
Best Local Similarity 93.8%; Pred. No. 5.4e-143;  
Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALITGYVYDHLTPLODMAHAG 75
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALITGYVYDHLTPLODMAHAG 242
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGSEIILGPADNPEGQGW 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGSEIILGPADNPEGQGW 302
QY 136 RLAPITAYSOQTRGLICITITSLTGRDKNOVEGEVQVSTATQSFPLATCVNGVCMVYFH 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 RLAPITAYSOQTRGLICITITSLTGRDKNOVEGEVQVSTATQSFPLATCVNGVCMVYFH 362
QY 196 GAGSKTLAGPKPIQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 GAGSKTLAGPKPIQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 422
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKAVDFIVESMET 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 423 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKAVDFIVESMET 482
QY 316 TMR 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 483 TMR 486

```

```

RESULT 11
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-11

```

```

Query Match 86.1%; Score 1525; DB 4; Length 1692;
Best Local Similarity 93.8%; Pred. No. 5.4e-143;
Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALITGYVYDHLTPLODMAHAG 75
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALITGYVYDHLTPLODMAHAG 242
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGSEIILGPADNPEGQGW 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGSEIILGPADNPEGQGW 302
QY 136 RLAPITAYSOQTRGLICITITSLTGRDKNOVEGEVQVSTATQSFPLATCVNGVCMVYFH 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 RLAPITAYSOQTRGLICITITSLTGRDKNOVEGEVQVSTATQSFPLATCVNGVCMVYFH 362
QY 196 GAGSKTLAGPKPIQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 GAGSKTLAGPKPIQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 422
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKAVDFIVESMET 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 423 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKAVDFIVESMET 482
QY 316 TMR 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 483 TMR 486

```

```

RESULT 12
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-9

```

```

Query Match 86.1%; Score 1525; DB 3; Length 2307;
Best Local Similarity 93.8%; Pred. No. 8.5e-143;
Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALITGYVYDHLTPLODMAHAG 75
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 275 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALITGYVYDHLTPLODMAHAG 334
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGSEIILGPADNPEGQGW 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 335 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGSEIILGPADNPEGQGW 394
QY 136 RLAPITAYSOQTRGLICITITSLTGRDKNOVEGEVQVSTATQSFPLATCVNGVCMVYFH 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 395 RLAPITAYSOQTRGLICITITSLTGRDKNOVEGEVQVSTATQSFPLATCVNGVCMVYFH 454
QY 196 GAGSKTLAGPKPIQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 455 GAGSKTLAGPKPIQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 514
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKAVDFIVESMET 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 515 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAVGIFRAAVCTRGVAKAVDFIVESMET 574
QY 316 TMR 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 575 TMR 578

```

```

RESULT 13
US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02

```

PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-9

Query Match 86.1%; Score 1525; DB 4; Length 2307;  
Best Local Similarity 93.8%; Pred. No. 8.5e-143;  
Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
DB 275 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 334  
QY 76 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARRGREIILGPADNFEQGM 135  
DB 335 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARRGREIILGPADNFEQGM 394  
QY 136 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTATOSFLATCVNGVCTVYH 195  
DB 395 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTATOSFLATCVNGVCTVYH 454  
QY 196 GAGSKTLAAGPKPITOMTNTVDODLVGMQAPPGARSITPCTCGSSDLVYTRHADVIYVR 255  
DB 455 GAGSKTLAAGPKPITOMTNTVDODLVGMQAPPGARSITPCTCGSSDLVYTRHADVIYVR 514  
QY 256 RRGDSRGSLSPPRVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315  
DB 515 RRGDSRGSLSPPRVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 574  
QY 316 TMR 319  
DB 575 TMR 578

RESULT 14  
US-09-539-601-33  
Sequence 33, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartschlagel, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 33  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-33

Query Match 86.1%; Score 1524; DB 4; Length 3010;  
Best Local Similarity 94.4%; Pred. No. 1.6e-142;  
Matches 287; Conservative 9; Mismatches 8; Indels 0; Gaps 0;  
QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
DB 904 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 963  
QY 76 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARRGREIILGPADNFEQGM 135

DB 964 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARRGREIILGPADNFEQGM 1023  
QY 136 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTATOSFLATCVNGVCTVYH 195  
DB 1024 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTATOSFLATCVNGVCTVYH 1083  
QY 196 GAGSKTLAAGPKPITOMTNTVDODLVGMQAPPGARSITPCTCGSSDLVYTRHADVIYVR 255  
DB 1084 GAGSKTLAAGPKPITOMTNTVDODLVGMQAPPGARSITPCTCGSSDLVYTRHADVIYVR 1143  
QY 256 RRGDSRGSLSPPRVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315  
DB 1144 RRGDSRGSLSPPRVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
QY 316 TMR 319  
DB 1204 TMR 1207

RESULT 15  
US-09-263-933-18  
Sequence 18, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 18  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-18

Query Match 85.6%; Score 1516; DB 3; Length 1692;  
Best Local Similarity 93.4%; Pred. No. 4.3e-142;  
Matches 284; Conservative 13; Mismatches 7; Indels 0; Gaps 0;  
QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
DB 183 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 242  
QY 76 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARRGREIILGPADNFEQGM 135  
DB 243 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARRGREIILGPADNFEQGM 302  
QY 136 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTATOSFLATCVNGVCTVYH 195  
DB 303 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTATOSFLATCVNGVCTVYH 362  
QY 196 GAGSKTLAAGPKPITOMTNTVDODLVGMQAPPGARSITPCTCGSSDLVYTRHADVIYVR 255  
DB 363 GAGSKTLAAGPKPITOMTNTVDODLVGMQAPPGARSITPCTCGSSDLVYTRHADVIYVR 422  
QY 256 RRGDSRGSLSPPRVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315  
DB 423 RRGDSRGSLSPPRVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 482  
QY 316 TMR 319  
DB 483 TMR 486

Search completed: May 6, 2004, 09:39:00  
Job time: 14.109 secs

Fri May 7 13:37:08 2004

us-10-650-585-4.rai

---

Page 7







Db 181 FLATCVNGVCMVTFHAGSGKTLGAPKGPITQMTTNDQDVLGWAQAPPGASMTPTCTGSS 240  
 QY 241 DLYLVRHADVIPIVRRRSGDSRGLSPVSYLKSGSGGDLCPSGHAGIFRAAVCTRG 300  
 Db 241 DLYLVRHADVIPIVRRRSGDSRGLSPVSYLKSGSGGDLCPSGHAGIFRAAVCTRG 300  
 QY 301 VAKAVDFIPVESMETTWTSSAMRHPOFGKXXX 334  
 Db 301 VAKAVDFIPVESMETTWTSSAMRHPOFGKXXX 334

## RESULT 2

US-10-650-585-4  
 ; Sequence 4, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; PRIOR FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 334  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-4

Query Match 100.0%; Score 1771; DB 16; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-163;  
 Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K K K K L E H H H H H S A G I T K V P F V R A G L I R A C M L V R K A G H Y V Q M A F M K L A A L T G T Y 60  
 Db 1 M K K K K L E H H H H H S A G I T K V P F V R A G L I R A C M L V R K A G H Y V Q M A F M K L A A L T G T Y 60  
 QY 61 V Y D H L T P L Q D W A H A G L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R 120  
 Db 61 V Y D H L T P L Q D W A H A G L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R 120  
 QY 121 E I L G P A D N F E G Q G R L A P I T A V S Q O T R G L G C I I T S L T R D K N O V E G E V Q V S T A T O S 180  
 Db 121 E I L G P A D N F E G Q G R L A P I T A V S Q O T R G L G C I I T S L T R D K N O V E G E V Q V S T A T O S 180  
 QY 181 F L A T C V N G V C M T V F H A G S K T L A G K G P I T Q M T T N D Q D L V G W A P P G A S M T P C T G S S 240  
 Db 181 F L A T C V N G V C M T V F H A G S K T L A G K G P I T Q M T T N D Q D L V G W A P P G A S M T P C T G S S 240  
 QY 241 D L Y L V R H A D V I P V R R R S G D S R G L S P R P V S Y L K S G S G G D L C P S G H A G I F R A A V C T R G 300  
 Db 241 D L Y L V R H A D V I P V R R R S G D S R G L S P R P V S Y L K S G S G G D L C P S G H A G I F R A A V C T R G 300  
 QY 301 V A K A V D F I P V E S M E T T W T S S A M R H P O F G K K K X 334  
 Db 301 V A K A V D F I P V E S M E T T W T S S A M R H P O F G K K K X 334

## RESULT 3

US-10-017-736-2  
 ; Sequence 2, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 409  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-017-736-2

Query Match 93.7%; Score 1660; DB 13; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 6e-152;  
 Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 A G I T K V P F V R A G L I R A C M L V R K A G H Y V Q M A F M K L A A L T G T Y V D H L T P L Q D W A H A G 75  
 Db 95 A G I T K V P F V R A G L I R A C M L V R K A G H Y V Q M A F M K L A A L T G T Y V D H L T P L Q D W A H A G 154  
 QY 76 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R R G E I I L G P A D N F E G Q G W 135  
 Db 155 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R R G E I I L G P A D N F E G Q G W 214  
 QY 136 R L L A P I T A V S Q O T R G L G C I I T S L T R D K N O V E G E V Q V S T A T O S F L A T C V N G V C M T V F H 195  
 Db 215 R L L A P I T A V S Q O T R G L G C I I T S L T R D K N O V E G E V Q V S T A T O S F L A T C V N G V C M T V F H 274  
 QY 196 G A G S K T L A G K G P I T Q M T T N D Q D L V G W A P P G A S M T P C T G S S D L Y L V R H A D V I P V R 255  
 Db 275 G A G S K T L A G K G P I T Q M T T N D Q D L V G W A P P G A S M T P C T G S S D L Y L V R H A D V I P V R 334  
 QY 256 R G G D S R G L S P R P V S Y L K S G S G G D L C P S G H A G I F R A A V C T R G V A K A V D F I P V E S M E T 315  
 Db 335 R G G D S R G L S P R P V S Y L K S G S G G D L C P S G H A G I F R A A V C T R G V A K A V D F I P V E S M E T 394  
 QY 316 T W R T S S A M R H P O F G K 330  
 Db 395 T W R T S S A M R H P O F G K 409

## RESULT 4

US-10-650-585-2  
 ; Sequence 2, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 409  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-2

Query Match 93.7%; Score 1660; DB 16; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 6e-152;  
 Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 A G I T K V P F V R A G L I R A C M L V R K A G H Y V Q M A F M K L A A L T G T Y V D H L T P L Q D W A H A G 75  
 Db 95 A G I T K V P F V R A G L I R A C M L V R K A G H Y V Q M A F M K L A A L T G T Y V D H L T P L Q D W A H A G 154  
 QY 76 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R R G E I I L G P A D N F E G Q G W 135  
 Db 155 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R R G E I I L G P A D N F E G Q G W 214  
 QY 136 R L L A P I T A V S Q O T R G L G C I I T S L T R D K N O V E G E V Q V S T A T O S F L A T C V N G V C M T V F H 195

```

Db 215 RLAPITAYSOQTGGLGCIITSLTGRDKQVEEVEVAVSTATQSFATCNGVCWTFVH 274
QY 196 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSMPTCTCGSSDLYLVTREADVIVR 255
Db 275 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSMPTCTCGSSDLYLVTREADVIVR 334
QY 256 RRGDSRGSLLSPREVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 335 RRGDSRGSLLSPREVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 394
QY 316 TMRSSAMRHPOFGG 330
Db 395 TMRSSAMRHPOFGG 409

```

## RESULT 5

```

US-10-017-736-10
; Sequence 10, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 303
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-10

```

```

Query Match 89.7%; Score 1589; DB 13; Length 303;
Best Local Similarity 100.0%; Pred. No. 3e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 16 AGITKYFVFRAGQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 1 AGITKYFVFRAGQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135
Db 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 120
QY 136 RLAPITAYSOQTGGLGCIITSLTGRDKQVEEVEVAVSTATQSFATCNGVCWTFVH 195
Db 121 RLAPITAYSOQTGGLGCIITSLTGRDKQVEEVEVAVSTATQSFATCNGVCWTFVH 180
QY 196 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSMPTCTCGSSDLYLVTREADVIVR 255
Db 181 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSMPTCTCGSSDLYLVTREADVIVR 240
QY 256 RRGDSRGSLLSPREVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 241 RRGDSRGSLLSPREVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
QY 316 TMR 318
Db 301 TMR 303

```

## RESULT 6

```

US-10-650-585-10
; Sequence 10, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082

```

```

; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 303
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-10

```

```

Query Match 89.7%; Score 1589; DB 16; Length 303;
Best Local Similarity 100.0%; Pred. No. 3e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 16 AGITKYFVFRAGQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 1 AGITKYFVFRAGQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135
Db 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 120
QY 136 RLAPITAYSOQTGGLGCIITSLTGRDKQVEEVEVAVSTATQSFATCNGVCWTFVH 195
Db 121 RLAPITAYSOQTGGLGCIITSLTGRDKQVEEVEVAVSTATQSFATCNGVCWTFVH 180
QY 196 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSMPTCTCGSSDLYLVTREADVIVR 255
Db 181 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSMPTCTCGSSDLYLVTREADVIVR 240
QY 256 RRGDSRGSLLSPREVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 241 RRGDSRGSLLSPREVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
QY 316 TMR 318
Db 301 TMR 303

```

## RESULT 7

```

US-10-017-736-14
; Sequence 14, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-14

```

```

Query Match 89.7%; Score 1589; DB 13; Length 341;
Best Local Similarity 100.0%; Pred. No. 3.5e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 16 AGITKYFVFRAGQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 39 AGITKYFVFRAGQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 98
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135

```

Db 99 LRLDAVAEPIVFSDMVEVKIITWGADTAACGDIISGLPVSARRGREITLLGPADNFEQGW 158  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGVVSTATQSLATCVNGVCTVPH 195  
 Db 159 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGVVSTATQSLATCVNGVCTVPH 218  
 QY 196 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 255  
 Db 219 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 278  
 QY 256 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 Db 279 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 338  
 QY 316 TMR 318  
 Db 339 TMR 341

RESULT 8  
 US-10-650-585-14  
 ; Sequence 14, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OR INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14  
 ; LENGTH: 341  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-14

Query Match 89.7%; Score 1589; DB 16; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-145;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 Db 39 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 98  
 QY 76 LRLDAVAEPIVFSDMVEVKIITWGADTAACGDIISGLPVSARRGREITLLGPADNFEQGW 135  
 Db 99 LRLDAVAEPIVFSDMVEVKIITWGADTAACGDIISGLPVSARRGREITLLGPADNFEQGW 158  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGVVSTATQSLATCVNGVCTVPH 195  
 Db 159 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGVVSTATQSLATCVNGVCTVPH 218  
 QY 196 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 255  
 Db 219 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 278  
 QY 256 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 Db 279 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 338  
 QY 316 TMR 318  
 Db 339 TMR 341

RESULT 9  
 US-10-017-736-13  
 ; Sequence 13, Application US/10017736

Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OR INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-017-736-13

Query Match 89.7%; Score 1589; DB 13; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-145;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 Db 50 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 109  
 QY 76 LRLDAVAEPIVFSDMVEVKIITWGADTAACGDIISGLPVSARRGREITLLGPADNFEQGW 135  
 Db 110 LRLDAVAEPIVFSDMVEVKIITWGADTAACGDIISGLPVSARRGREITLLGPADNFEQGW 169  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGVVSTATQSLATCVNGVCTVPH 195  
 Db 170 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGVVSTATQSLATCVNGVCTVPH 229  
 QY 196 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 255  
 Db 230 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 289  
 QY 256 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 Db 290 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 349  
 QY 316 TMR 318  
 Db 350 TMR 352

RESULT 10  
 US-10-650-585-13  
 ; Sequence 13, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OR INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-13

Query Match 89.7%; Score 1589; DB 16; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-145;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 75

```

Db 50 AGITKVPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 109
Qy 76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPDNFEQGW 135
Db 110 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPDNFEQGW 169
Qy 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVPH 195
Db 170 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVPH 229
Qy 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTTRADVI 255
Db 230 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTTRADVI 289
Qy 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 290 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 349
Qy 316 TMR 318
Db 350 TMR 352

```

```

RESULT 11
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-12

```

```

Query Match 89.7%; Score 1589; DB 13; Length 380;
Best Local Similarity 100.0%; Pred. No. 4.1e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 16 AGITKVPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 78 AGITKVPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 137
Qy 76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPDNFEQGW 135
Db 138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPDNFEQGW 197
Qy 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVPH 195
Db 198 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVPH 257
Qy 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTTRADVI 255
Db 258 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTTRADVI 317
Qy 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 318 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 377
Qy 316 TMR 318
Db 378 TMR 380

```

```

RESULT 12
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-12

```

```

Query Match 89.7%; Score 1589; DB 16; Length 380;
Best Local Similarity 100.0%; Pred. No. 4.1e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 16 AGITKVPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 78 AGITKVPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 137
Qy 76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPDNFEQGW 135
Db 138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPDNFEQGW 197
Qy 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVPH 195
Db 198 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVPH 257
Qy 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTTRADVI 255
Db 258 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTTRADVI 317
Qy 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 318 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 377
Qy 316 TMR 318
Db 378 TMR 380

```

```

RESULT 13
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

```

```

Query Match 89.7%; Score 1589; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.3e-145;

```

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLLTPLODMAHAG 75  
 DB 91 AGITKVPYFVRAOGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLLTPLODMAHAG 150  
 QY 76 LRDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135  
 DB 151 LRDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 210  
 QY 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVFH 195  
 DB 211 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVFH 270  
 QY 196 GAGSKTLAAGPKGPIITOMYTNVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIPIVR 255  
 DB 271 GAGSKTLAAGPKGPIITOMYTNVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIPIVR 330  
 QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 331 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 390  
 QY 316 TMR 318  
 DB 391 TMR 393

## RESULT 14

US-10-650-585-11  
 ; Sequence 11, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 11  
 ; LENGTH: 393  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-11

Query Match 89.7%; Score 1589; DB 16; Length 393;

Best Local Similarity 100.0%; Pred. No. 4,3e-145; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLLTPLODMAHAG 75  
 DB 91 AGITKVPYFVRAOGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLLTPLODMAHAG 150  
 QY 76 LRDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135  
 DB 151 LRDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 210  
 QY 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVFH 195  
 DB 211 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVFH 270  
 QY 196 GAGSKTLAAGPKGPIITOMYTNVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIPIVR 255  
 DB 271 GAGSKTLAAGPKGPIITOMYTNVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIPIVR 330  
 QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 331 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

QY 316 TMR 318  
 DB 391 TMR 393

## RESULT 15

US-10-017-736-18  
 ; Sequence 18, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 18  
 ; LENGTH: 303  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-017-736-18

Query Match 89.2%; Score 1580; DB 13; Length 303;

Best Local Similarity 99.7%; Pred. No. 2.2e-144; Mismatches 1; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLLTPLODMAHAG 75  
 DB 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLLTPLODMAHAG 60  
 QY 76 LRDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135  
 DB 61 LRDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 120  
 QY 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVFH 195  
 DB 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVFH 180  
 QY 196 GAGSKTLAAGPKGPIITOMYTNVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIPIVR 255  
 DB 181 GAGSKTLAAGPKGPIITOMYTNVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIPIVR 240  
 QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 241 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 QY 316 TMR 318  
 DB 301 TMR 303

Search completed: May 6, 2004, 09:43:18  
 Job time : 36.167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 41.2434 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-10

Sequence: 1 AGITKVPYFRAQLIRACM.....RGYAKAVDPFVPSMETTMR 303

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1589	100.0	303	ABG32183	ABG32183 HCV prote
2	1589	100.0	334	ABG32182	ABG32182 HCV prote
3	1589	100.0	341	ABG32187	ABG32187 HCV prote
4	1589	100.0	352	ABG32186	ABG32186 HCV prote
5	1589	100.0	380	ABG32185	ABG32185 HCV prote
6	1589	100.0	393	ABG32184	ABG32184 HCV prote
7	1589	100.0	409	ABG32181	ABG32181 HCV prote
8	1580	99.4	303	ABG32191	ABG32191 HCV prote
9	1579	99.4	303	ABG32189	ABG32189 HCV prote
10	1570	98.8	301	ABG32190	ABG32190 HCV prote
11	1532	96.4	292	ABG32188	ABG32188 HCV prote
12	1530	96.3	2201	ABG30601	ABG30601 Hepatitis
13	1530	96.3	2201	ABG30591	ABG30591 Hepatitis
14	1530	96.3	2201	ABG30600	ABG30600 Hepatitis
15	1530	96.3	2201	ABG30581	ABG30581 Hepatitis
16	1530	96.3	2201	ABG30593	ABG30593 Hepatitis
17	1530	96.3	2201	ABG30582	ABG30582 Hepatitis
18	1530	96.3	2201	ABG30580	ABG30580 Hepatitis
19	1530	96.3	2201	ABG30587	ABG30587 Hepatitis
20	1530	96.3	2201	ABG30599	ABG30599 Hepatitis
21	1530	96.3	2201	ABG30594	ABG30594 Hepatitis
22	1530	96.3	2201	ABG30598	ABG30598 Hepatitis
23	1530	96.3	2201	ABG30595	ABG30595 Hepatitis
24	1530	96.3	3010	ABG32458	ABG32458 Hepatitis
25	1530	96.3	3010	ABG32459	ABG32459 Hepatitis

26	1530	96.3	3010	5	ABG32451	Abg32451 Hepatitis
27	1530	96.3	3010	5	ABG32455	Abg32455 Hepatitis
28	1530	96.3	3010	5	ABG32457	Abg32457 Hepatitis
29	1530	96.3	3010	5	ABG32460	Abg32460 Hepatitis
30	1530	96.3	3010	5	ABG32461	Abg32461 Hepatitis
31	1530	96.3	3010	5	ABG32454	Abg32454 Hepatitis
32	1530	96.3	3010	5	ABG32477	Abg32477 HCV-S1 fu
33	1530	96.3	3011	5	ABG32456	Abg32456 Hepatitis
34	1527	96.1	2201	5	ABG30586	Abg30586 Hepatitis
35	1527	96.1	2201	5	ABG30589	Abg30589 Hepatitis
36	1527	96.1	2201	5	ABG30583	Abg30583 Hepatitis
37	1527	96.1	2201	5	ABG30588	Abg30588 Hepatitis
38	1527	96.1	2307	3	AA770064	AA770064 Recombina
39	1527	96.1	3010	2	AA868622	AA868622 HCV prote
40	1527	96.1	3010	2	AA882694	AA882694 Partial H
41	1526	96.0	2201	5	ABG30590	Abg30590 Hepatitis
42	1524	95.9	2307	3	AA770065	AA770065 Recombina
43	1524	95.9	3010	5	ABG32452	Abg32452 Hepatitis
44	1523	95.8	2201	5	ABG30584	Abg30584 Hepatitis
45	1523	95.8	2201	5	ABG30602	Abg30602 Hepatitis

## ALIGNMENTS

RESULT 1  
ABG32183  
ID ABG32183 standard; protein; 303 AA.  
AC ABG32183;  
XX  
XX  
DT 05-NOV-2002 (first entry)  
XX  
XX HCV protease NS2/3 truncation mutant 904-1206.  
DE  
XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
KW chronic liver disease; cirrhosis; end-stage liver disease; virologic;  
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KW chaotropic agent; mutant; mutain.  
XX  
XX Hepatitis C virus.  
OS  
OS Synthetic.  
XX  
XX WO200248375-A2.  
XX  
XX 20-JUN-2002.  
XX  
XX 13-DEC-2001; 2001WO-CA001796.  
XX  
XX 15-DEC-2000; 2000US-0256031P.  
XX  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.  
XX  
XX Claim 39; Page 58-59; 67pp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
XX its truncation or a mutated sequence, where the protease is in a solution  
XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide



CC appearing as ABG32182; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC chaperone, involving isolating the protease in the presence of a  
 CC chaperone agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LMO in the presence of reduced concentration of the  
 CC chaperone agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded, inactive NS2/3 protease in a medium  
 CC containing an activating detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 904-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 303 AA;

Query Match 100.0%; Score 1589; DB 5; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVPAQGLIRACMLVRRKAGGHVQMAFMKLAALTGYVYDHLTPLODMAHAG 60  
 DB 1 AGITKVPYFVPAQGLIRACMLVRRKAGGHVQMAFMKLAALTGYVYDHLTPLODMAHAG 60  
 QY 61 LRDLAVAVEPIFSDMEVKIITWGADTPAACGDIISGIPVARSRRREILGPADEFGQGM 120  
 DB 61 LRDLAVAVEPIFSDMEVKIITWGADTPAACGDIISGIPVARSRRREILGPADEFGQGM 120  
 QY 121 RLAPITAYSQQTRGGLGCTITSTGERDKVQGEVQVNSITAFOSFATCVNVCMTVH 180  
 DB 121 RLAPITAYSQQTRGGLGCTITSTGERDKVQGEVQVNSITAFOSFATCVNVCMTVH 180  
 QY 181 GAGSKTAGPKGPIITQMTVNDQDLVGMQAPFGARSWTPTCCSSDLYLTRADYIPVR 240  
 DB 181 GAGSKTAGPKGPIITQMTVNDQDLVGMQAPFGARSWTPTCCSSDLYLTRADYIPVR 240  
 QY 241 RRGDSRGSLLSPRVSYLKSGSGPFLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMET 300  
 DB 241 RRGDSRGSLLSPRVSYLKSGSGPFLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMET 300  
 QY 301 TMR 303  
 DB 301 TMR 303

RESULT 2  
 ABG32182  
 ID ABG32182 standard; protein; 334 AA.

XX AC ABG32182;  
 XX DT 05-NOV-2002 (first entry)  
 XX DE HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.  
 XX KM HCV, enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
 XX KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 XX KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX KM chaotropic agent; 4K-6H (904-1206)st-4K; mutant; muten.  
 XX OS Hepatitis C virus.  
 XX XX Synthetic.

EH Key Location/Qualifiers  
 FT Peptide 1..15  
 FT Protein /note="4-Lys/His tag"  
 FT Peptide 16..302  
 FT Peptide /note="Truncated NS2/3 protease"  
 FT Peptide 319..334  
 FT Peptide /note="Streptavidin/4-Lys tag"  
 XX W0200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOHR) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibault D, Lamare D, Maurice R, Plote L, Pause A;  
 XX WPI, 2002-599511/64.  
 XX N-PSDB; ABR90407.  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 XX useful as therapeutic agents against hepatitis C virus, comprises full  
 XX length non-structural protease, or its truncation.  
 XX Claim 39; Fig 9B; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32182; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaperone agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LMO in the presence of reduced concentration of the  
 CC chaperone agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activating detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of.  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 4K-6H (904-  
 CC 1206)st-4K comprising a truncated NS2/3 protein with a four lys/six His N  
 CC -terminal tag, a C-terminal streptavidin tag and C-terminal four lys tag

XX Sequence 334 AA;  
 XX Query Match 100.0%; Score 1589; DB 5; Length 334;  
 XX Best Local Similarity 100.0%; Pred. No. 5.1e-146;  
 XX Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVPAQGLIRACMLVRRKAGGHVQMAFMKLAALTGYVYDHLTPLODMAHAG 60  
 DB 1 AGITKVPYFVPAQGLIRACMLVRRKAGGHVQMAFMKLAALTGYVYDHLTPLODMAHAG 75

QY 61 LRLDAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 120  
 Db 76 LRLDAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135  
 QY 121 RLAPITAYSQQTRGLGCIITSLTGRDKNQVEGVVSTATQSFPLATCVNVCWTFVH 180  
 Db 136 RLAPITAYSQQTRGLGCIITSLTGRDKNQVEGVVSTATQSFPLATCVNVCWTFVH 195  
 QY 181 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVR 240  
 Db 196 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVR 255  
 QY 241 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 256 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 QY 301 TMR 303  
 Db 316 TMR 318

## RESULT 3

ABG32187  
 ID ABG32187 standard; protein; 341 AA.

AC ABG32187;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 866-1206.

XX HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;

KM chronic liver disease; cirrhosis; end-stage liver disease; virucide;

KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;

KM chaotropic agent; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

PN WO200248375-A2.

XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CA001796.

XX 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases

XX useful as therapeutic agents against hepatitis C virus, comprises full

XX length non-structural protease, or its truncation.

XX Claim 41; Page 62-63; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminus  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a

CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 866-1206  
 CC (numbered relative to the full length NS2/3 protein)

SO Sequence 341 AA;

Query Match 100.0%; Score 1589; DB 5; Length 341;

Best Local Similarity 100.0%; Pred. No. 6e-146;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALITGVYVYDHLPLQDMAHAG 60  
 Db 39 AGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALITGVYVYDHLPLQDMAHAG 98  
 QY 61 LRLDAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 120  
 Db 99 LRLDAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 158  
 QY 121 RLAPITAYSQQTRGLGCIITSLTGRDKNQVEGVVSTATQSFPLATCVNVCWTFVH 180  
 Db 159 RLAPITAYSQQTRGLGCIITSLTGRDKNQVEGVVSTATQSFPLATCVNVCWTFVH 218  
 QY 181 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVR 240  
 Db 219 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVR 278  
 QY 241 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 279 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 338  
 QY 301 TMR 303  
 Db 339 TMR 341

RESULT 4  
 ABG32186  
 ID ABG32186 standard; protein; 352 AA.  
 XX ABG32186;  
 AC  
 DT 05-NOV-2002 (first entry)  
 DE HCV protease NS2/3 truncation mutant 855-1206.  
 XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; virucide;  
 KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 PN WO200248375-A2.  
 XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX WPI; 2002-599511/64.  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 41; Page 61-62; 67pp; English.  
 XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC present sequence represents the NS2/3 truncation mutant 855-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 CC XX  
 CC Sequence 352 AA;  
 CC  
 CC Query Match 100.0%; Score 1589; DB 5; Length 352;  
 CC Best Local Similarity 100.0%; Pred. No. 6, 3e-146;  
 CC Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 1 AGITKVYFVRAGLIRACMLVRKAGHYVQAFMKALATITTYYYDHLPIQDPAHAG 60  
 CC 50 AGITKVYFVRAGLIRACMLVRKAGHYVQAFMKALATITTYYYDHLPIQDPAHAG 109  
 CC 61 LRDLAANVEVITSDMEVKIITGADTACGDIISGLPVASRGREILGPNPFGQGM 120  
 CC 110 LRDLAANVEVITSDMEVKIITGADTACGDIISGLPVASRGREILGPNPFGQGM 169  
 CC 121 RLAPITAYSQOTRGILGCIITISLGRDKNOVEGEVQVSTATQSFPLATCVNGCVCTVPH 180  
 CC 170 RLAPITAYSQOTRGILGCIITISLGRDKNOVEGEVQVSTATQSFPLATCVNGCVCTVPH 229  
 CC 181 GAGSKTAGRGKPTTQWNTVDDIVGMOAPPARASMTPTCGSSDLVYLRADYIPVR 240  
 CC 230 GAGSKTAGRGKPTTQWNTVDDIVGMOAPPARASMTPTCGSSDLVYLRADYIPVR 289

QY 241 RRGDSRGLSPREVSYLKGSGGFLPCSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 290 RRGDSRGLSPREVSYLKGSGGFLPCSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 349  
 QY 301 TMR 303  
 DB 350 TMR 352  
 RESULT 5  
 ID ABG32185 standard; protein, 380 AA.  
 AC ABG32185;  
 XX 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 truncation mutant 827-1206.  
 XX HCV, enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; viruslike;  
 XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutcin.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 PN WO200248375-A2.  
 PD 20-JUN-2002.  
 PF 13-DEC-2001; 2001WO-CA001796.  
 PR 15-DEC-2000; 2000US-0256031P.  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 PI WPI; 2002-599511/64.  
 DR Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 41; Page 60-61; 67pp; English.  
 XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is

CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation mutant 827-1206  
CC (numbered relative to the full length NS2/3 protein)

XX Sequence 380 AA;

Query Match 100.0%; Score 1589; DB 5; Length 380;

Best Local Similarity 100.0%; Pred. No. 7e-146; Mismatches 0; Gaps 0;

Matches 303; Conservative 0; Indels 0; Gaps 0;

QY 1 AGITVYFVFAOGLIRACMLVRKAGGHVYQMAFMKLAALTGTYVDHLTPLODMAHAG 60  
DB 78 AGITVYFVFAOGLIRACMLVRKAGGHVYQMAFMKLAALTGTYVDHLTPLODMAHAG 137  
QY 61 LRDLAAVEPVIFSDMEVKIITWGDPTAACGDIISGLPVASARRREILLGPADNFEQGW 120  
DB 138 LRDLAAVEPVIFSDMEVKIITWGDPTAACGDIISGLPVASARRREILLGPADNFEQGW 197  
QY 121 RLAPITAVSQOTRGLICITITSLGRDNQVEGVEGVVSTATOSFLATCVNGVCTVFFH 180  
DB 198 RLAPITAVSQOTRGLICITITSLGRDNQVEGVEGVVSTATOSFLATCVNGVCTVFFH 257  
QY 181 GAGSKTLAGKPGITQWYTNVDQLVGQAPPGARSMPTCTGSSDLVLRHADVIPIVR 240  
DB 258 GAGSKTLAGKPGITQWYTNVDQLVGQAPPGARSMPTCTGSSDLVLRHADVIPIVR 317  
QY 241 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
DB 318 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 377  
QY 301 TMR 303  
DB 378 TMR 360

RESULT 6

ABG32184

ID ABG32184 standard; protein; 393 AA.

XX ABG32184;

AC 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 815-1206.

XX HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
XX hepatocellular carcinoma; anti-inflammatory; lauryldiethyamine oxide; LDAO;  
XX chaotropic agent; mutant; mutagen.

OS Hepatitis C virus.  
XX Synthetic.

PN WO200248375-A2.

PD 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CA001796.

PR 15-DEC-2000; 2000US-0256031P.

XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Ttibbeault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-559511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full

PT length non-structural protease, or its truncation.

XX Claim 41; Page 59-60; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethyamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation mutant 815-1206  
CC (numbered relative to the full length NS2/3 protein)

XX Sequence 393 AA;

Query Match 100.0%; Score 1589; DB 5; Length 393;

Best Local Similarity 100.0%; Pred. No. 7.3e-146; Mismatches 0; Indels 0; Gaps 0;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITVYFVFAOGLIRACMLVRKAGGHVYQMAFMKLAALTGTYVDHLTPLODMAHAG 60  
DB 91 AGITVYFVFAOGLIRACMLVRKAGGHVYQMAFMKLAALTGTYVDHLTPLODMAHAG 150  
QY 61 LRDLAAVEPVIFSDMEVKIITWGDPTAACGDIISGLPVASARRREILLGPADNFEQGW 120  
DB 151 LRDLAAVEPVIFSDMEVKIITWGDPTAACGDIISGLPVASARRREILLGPADNFEQGW 210  
QY 121 RLAPITAVSQOTRGLICITITSLGRDNQVEGVEGVVSTATOSFLATCVNGVCTVFFH 180  
DB 211 RLAPITAVSQOTRGLICITITSLGRDNQVEGVEGVVSTATOSFLATCVNGVCTVFFH 270  
QY 181 GAGSKTLAGKPGITQWYTNVDQLVGQAPPGARSMPTCTGSSDLVLRHADVIPIVR 240  
DB 271 GAGSKTLAGKPGITQWYTNVDQLVGQAPPGARSMPTCTGSSDLVLRHADVIPIVR 330  
QY 241 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
DB 331 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 390  
QY 301 TMR 303  
DB 391 TMR 393

RESULT 7

ABG32181

ID ABG32181 standard; protein; 409 AA.

XX

AC ABG32181;  
 DT 05-NOV-2002 (first entry)  
 DE HCV protease NS2/3 (810-1206).  
 XX HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 KM hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT Peptide 398..409  
 FT /note="Streptavidin tag"  
 XX MO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001MO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 PI WPI; 2002-599511/64.  
 DR N-PSDB; ABK90406.  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 42; Fig 1B; 67pp; English.  
 XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) ID-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 (810-1206) protein, which has a C-  
 CC terminal streptavidin tag

SQ Sequence 409 AA;  
 Query Match 100.0%; Score 1589; DB 5; Length 409;  
 Best local similarity 100.0%; Pred. No. 7,8e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGITKVFYFVRAQGLIRACMLVRKAGHYVQVAFMKALATGTYVYDHLTPLOMAHAG 60  
 DB 95 AGITKVFYFVRAQGLIRACMLVRKAGHYVQVAFMKALATGTYVYDHLTPLOMAHAG 154  
 QY 61 LRDLAVAVEPITPSDMVKIITWQADTAACGDIISGLPVARRGREILLGPANFEGQGW 120  
 DB 155 LRDLAVAVEPITPSDMVKIITWQADTAACGDIISGLPVARRGREILLGPANFEGQGW 214  
 QY 121 RLAPITRAYSQOTGLGCIITSLTGRDKNQVEGVQVSTATQSFATCVCWCVTFH 180  
 DB 215 RLAPITRAYSQOTGLGCIITSLTGRDKNQVEGVQVSTATQSFATCVCWCVTFH 274  
 QY 181 GAGSKTLAGPKPITQWYTNVDQVLVGMQAPPGARSMTPCTCGSSDLVLYTRHADVTPVR 240  
 DB 275 GAGSKTLAGPKPITQWYTNVDQVLVGMQAPPGARSMTPCTCGSSDLVLYTRHADVTPVR 334  
 QY 241 RRGDSRGSILSPRVSVLYKSSGGPILCPGSHAVGIPRAVCTRGVAKADVFPVESMET 300  
 DB 335 RRGDSRGSILSPRVSVLYKSSGGPILCPGSHAVGIPRAVCTRGVAKADVFPVESMET 394  
 QY 301 TMR 303  
 DB 395 TMR 397  
 RESULT 8  
 ABG32191  
 ID ABG32191 standard; protein; 303 AA.  
 AC ABG32191;  
 XX 05-NOV-2002 (first entry)  
 DT HCV protease NS2/3 truncation 904-1206/Cys9393Aa.  
 DE HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 KM hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT Misc-difference 90 /note="Wild-type Cys substituted by Ala"  
 XX MO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001MO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 PI WPI; 2002-599511/64.  
 DR Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Disclosure; Page 65-66; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation of a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32189; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterization of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant Cys993Ala (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity

XX Sequence 303 AA;

Query Match 99.4%; Score 1580; DB 5; Length 303;  
Best Local Similarity 99.7%; Pred. No. 3.8e-145;  
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGITKVPYFRAOGLIPACMLVRKAGHYVNAEMKLAALTGYVDHITPLQDMAHAG 60  
DB 1 AGITKVPYFRAOGLIPACMLVRKAGHYVNAEMKLAALTGYVDHITPLQDMAHAG 60  
QY 61 LRDLAVALVEPVISFDMVKIITWGADTFAAGSDIISGLPVASRRGRILLGPANFEGGQM 120  
DB 61 LRDLAVALVEPVISFDMVKIITWGADTFAAGSDIISGLPVASRRGRILLGPANFEGGQM 120  
QY 121 RLAPITAYVQQQTRGLIGCIITSLTGEDKNQVGEVAVSTATQSFATCVNGVCTVPH 180  
DB 121 RLAPITAYVQQQTRGLIGCIITSLTGEDKNQVGEVAVSTATQSFATCVNGVCTVPH 180  
QY 181 GAGSKTLAGKSPITQWYTNVDDIVKQAPPGARSMTPCTCGSSDLVLTTRADVIPIR 240  
DB 181 GAGSKTLAGKSPITQWYTNVDDIVKQAPPGARSMTPCTCGSSDLVLTTRADVIPIR 240  
QY 241 RRDSRGSLLSPRPVSTLKGSSGGLLCPGHAIVGIFRAVCTRGYAKAVDFIPVSMKT 300  
DB 241 RRDSRGSLLSPRPVSTLKGSSGGLLCPGHAIVGIFRAVCTRGYAKAVDFIPVSMKT 300  
QY 301 TMR 303  
DB 301 TMR 303

RESULT 9  
ABG32189  
ID ABG32189 standard; protein; 303 AA.

XX AC ABG32189;  
XX DT 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation 904-1206/His952Ala.  
DE HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; muten.  
XX Hepatitis C virus.  
OS Synthetic.  
XX Key Location/Qualifiers  
FH Misc-difference 45 /note= "wild-type His substituted by Ala"  
FT WO200248375-A2.  
XX 20-JUN-2002.  
XX 13-DEC-2001; 2001WO-CA001796.  
XX 15-DEC-2000; 2000US-0256031P.  
XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.  
XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PI length non-structural protease, or its truncation.  
XX Example 7, Fig 8, 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation of a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32189; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterization of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant His952Ala (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity

XX Sequence 303 AA;

Query Match 99.4%; Score 1579; DB 5; Length 303;



Best Local Similarity 99.7%; Pred. No. 4.8e-145;  
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGITKPYFVFAOGLIRACMLVRKAGHYVQVAFMKLAALGTYYVDHLTPLODMAHAG 60  
DB 1 AGITKPYFVFAOGLIRACMLVRKAGHYVQVAFMKLAALGTYYVDHLTPLODMAHAG 60  
QY 61 LRLAIAVAEPVIFSDMEVKIITWGADTAACDIIISGLPVASRRREILLGPADNFEQGM 120  
DB 61 LRLAIAVAEPVIFSDMEVKIITWGADTAACDIIISGLPVASRRREILLGPADNFEQGM 120  
QY 121 RLIAPIITAVSQOTRGLGCIITSLTRGRDKNOVEGEVQVSTATQSPFLATCVNGVCTVH 180  
DB 121 RLIAPIITAVSQOTRGLGCIITSLTRGRDKNOVEGEVQVSTATQSPFLATCVNGVCTVH 180  
QY 181 GAGSKTLAGPKGPITQMTNTVDOLVGMQAPPGARSMTCTCGSSDLVLTSHADVIPIVR 240  
DB 181 GAGSKTLAGPKGPITQMTNTVDOLVGMQAPPGARSMTCTCGSSDLVLTSHADVIPIVR 240  
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVSEMET 300  
DB 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVSEMET 300  
QY 301 TMR 303  
DB 301 TMR 303

RESULT 10  
ABG32190  
ID ABG32190 standard; protein; 301 AA.

AC ABG32190;  
DT 05-NOV-2002 (first entry)  
DE HCV protease NS2/3 truncation 904-1206/deltaLeu1026-1A1027.

XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
KW chronic liver disease; cirrhosis; end-stage liver disease; virologic;  
RW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; mutein.

OS Hepatitis C virus.  
XX Synthetic.

XX Key Location/Qualifiers  
FH Misc-difference 122..123  
FT /note= "Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"

XX WO200248375-A2.

XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CM001796.

XX 15-DEC-2000; 2000US-0256031P.

XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.

XX Example 7; Page 64-65; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal

CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1p-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent, and LDAO in the presence of reduced concentration of the  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation 904-1206 mutant  
CC deltaLeu1026-1A1027 (numbered relative to the full length NS2/3 protein)  
CC a mutant devoid of autocatalytic activity

XX Sequence 301 AA;

XX Query Match 98.8%; Score 1570; DB 5; Length 301;  
XX Best Local Similarity 99.3%; Pred. No. 3.6e-144;  
XX Matches 301; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 AGITKPYFVFAOGLIRACMLVRKAGHYVQVAFMKLAALGTYYVDHLTPLODMAHAG 60  
DB 1 AGITKPYFVFAOGLIRACMLVRKAGHYVQVAFMKLAALGTYYVDHLTPLODMAHAG 60  
QY 61 LRLAIAVAEPVIFSDMEVKIITWGADTAACDIIISGLPVASRRREILLGPADNFEQGM 120  
DB 61 LRLAIAVAEPVIFSDMEVKIITWGADTAACDIIISGLPVASRRREILLGPADNFEQGM 120  
QY 121 RLIAPIITAVSQOTRGLGCIITSLTRGRDKNOVEGEVQVSTATQSPFLATCVNGVCTVH 180  
DB 121 RLIAPIITAVSQOTRGLGCIITSLTRGRDKNOVEGEVQVSTATQSPFLATCVNGVCTVH 180  
QY 181 GAGSKTLAGPKGPITQMTNTVDOLVGMQAPPGARSMTCTCGSSDLVLTSHADVIPIVR 240  
DB 181 GAGSKTLAGPKGPITQMTNTVDOLVGMQAPPGARSMTCTCGSSDLVLTSHADVIPIVR 240  
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVSEMET 300  
DB 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVSEMET 298  
QY 301 TMR 303  
DB 301 TMR 303

RESULT 11  
ABG32188  
ID ABG32188 standard; protein; 292 AA.

XX ABG32188;

XX 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 915-1206.



KW	HCV; enzyme; protease; NS2/3 (915-1206); hepatitis C virus infection;
KV	chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
KJ	hepatotropic; antiinflammatory; lauryldimethylamine oxide; LDAO;
KM	chaotropic agent; mutant; mutein.
XX	
OS	Hepatitis C virus.
OS	Synthetic.
PN	WO200248375-A2.
PX	
PP	20-JUN-2002.
PD	
PF	13-DEC-2001; 2001WO-COA01796.
XX	
PR	15-DEC-2000; 2000US-0256031P.
XX	
PA	(BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.
XX	
PI	Tibbeault D, Lemarrie D, Maurice R, Pilote J, Pause A;
DR	WPI, 2002-599511/64.
XX	
PT	Novel polypeptide for screening inhibitors of non-structural proteases
PT	useful as therapeutic agents against hepatitis C virus, comprises full
PI	length non-structural protease, or its truncation.
XX	
PS	Claim 41; Page 63; 67pp; English.
CC	The invention relates to an isolated polypeptide consisting of a full-
CC	length HCV (hepatitis C virus) non-structural (NS2/3) protease (referred
CC	to also as NS2/3 (610-1206)), or its truncation, having as its N-terminal
CC	residue amino acid 810 to 906, or having a minimal amino acid sequence
CC	from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC	NS2/3 protease. Also included are (1) a composition (C) comprising an
CC	isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC	its truncation or a mutated sequence, where the protease is in a solution
CC	comprising a sufficient concentration of lauryldimethylamine oxide (LDAO)
CC	to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC	appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC	protease, involving isolating the protease in the presence of a
CC	chaotropic agent, refolding the isolated protease by contacting it with a
CC	reducing agent, and LDAO in the presence of reduced concentration of the
CC	chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC	protease, involving diluting/refolding inactive NS2/3 protease in a medium
CC	containing an activation detergent to induce auto-cleavage of the NS2/3
CC	protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC	protease, involving incubating the active NS2/3 protease produced by M2
CC	for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC	cleavage products or their fragments, and measuring the presence or
CC	absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC	; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC	active NS2/3 protease, involving carrying out M3 in the presence of, or
CC	absence of the potential inhibitor, comparing the amount of uncleaved
CC	NS2/3 protease, cleavage products or their fragments. The protease is
CC	useful for detailed biochemical characterisation of the enzymes and in
CC	the development of in vitro assays for screening novel inhibitors of
CC	NS2/3 protease which are useful as therapeutic agents against HCV
CC	infection (which causes chronic liver disease, cirrhosis and end-stage
CC	liver disease. M1 is useful for high level production of protease. The
CC	present sequence represents the NS2/3 truncation mutant 915-1206
CC	(numbered relative to the full length NS2/3 protein)
XQ	
XX	Sequence 292 AA:
Query Match	96.4%; Score 1532, DB 5; Length 292;
Best Local Similarity	100.0%, Fred. No. 1.7e-140;
Matches 292; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	12 AGGLIRACMLVKKAAGHYQMAEFKALATIGTYVDHLFPLQDMARAGRDLAAVEPY 71
Dz	1 AQGLIRACMLVKKAAGHYQMAEFKALATIGTYVDHLFPLQDMARAGRDLAAVEPY 60
YY	72 IPSDWEKVIITMGADTAACGDIIISGLIPVSANRGEIILGPANDNEGGWFLLAPITVAEQ 131

Db	61	IFSMEMKILITWGDJTRACDDISGLPVARRRGRELLLPADNFEQGOMLLAPITAYSQ	120
Qy	132	QTRBLIGCITITSLTGRDKVQVEGEVQVSTATQSPFLATCNGVCWTFHAGSKTLGAPK	191
Db	121	QTRBLIGCITITSLTGRDKVQVEGEVQVSTATQSPFLATCNGVCWTFHAGSKTLGAPK	180
Qy	192	GPITQMTNTNDQDLVGMQAPPGASRSMPTCGSSDLVLTVRADVIPRRRSDRGSLLS	25
Db	181	GPITQMTNTNDQDLVGMQAPPGASRSMPTCGSSDLVLTVRADVIPRRRSDRGSLLS	240
Qy	252	PRPVSYIKGSSGGPPLCPGSHAVGIFRAAVCTRGVAKAVDPIPVSMETTWK	303
Db	241	PRPVSYIKGSSGGPPLCPGSHAVGIFRAAVCTRGVAKAVDPIPVSMETTWK	292
RESULT 12			
Id	ABG30601	standard; protein; 2201 AA.	
XX	ABG30601;		
XX	AC		
XX	DT	21-OCT-2002 (first entry)	
DE	Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.		
XX	Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;		
KW	cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.		
XX	Hepatitis C virus.		
OS	Synthetic.		
XX	Key	Location/Qualifiers	
FT	Misc-difference	882	
FT	/label= Arg, Lys		
FT	Misc-difference	2183	
FT	/note= "Wild type Met substituted by Thr"		
XX	WO200252015-A2.		
PN	04-JUL-2002.		
XX	20-DEC-2001; 2001WO-CA001843.		
PF	22-DEC-2000; 2000US-0257857P.		
FR	(BOEH ) BOEHRINGER INGELHEIM CANADA LTD.		
XX	Kukolj G, Pause A;		
PI	WPI; 2002-575382/61.		
PT	New self-replicating RNA molecules from Hepatitis C virus (HCV), which		
PT	possess enhanced transduction or replication efficiency, useful for		
PT	evaluating potential inhibitors of HCV replication.		
XX	Claim 3; Page; 140pp; English.		
PS	The invention describes a self-replicating hepatitis C virus (HCV)		
CC	polynucleotide molecule comprising a 5'-non translated region (NTR),		
CC	where guanine at position 1 is substituted for adenine, a HCV polypeptide		
CC	region coding for a HCV polypeptide; and a 3'-NTR region. The self-		
CC	replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating		
CC	potential inhibitors of HCV replication. The HCV RNA molecule is also		
CC	useful for efficiently establishing cell culture replication. The self-		
CC	replicating polynucleotide molecule contains a 5'-NTR, where G at		
CC	position 1 is substituted for A, and therefore provides an alternative to		
CC	existing systems comprising a self-replicating HCV RNA molecule that, in		
CC	conjunction with mutations in the HCV non-structural region, such as the		
CC	G12042/C/R mutations, transduces and/or replicates with greater		
CC	efficiency. This amino acid sequence represents a mutant of the hepatitis		
CC	C virus replicon ApAxi2 and contains the viral protease NS2/3, protease		
CC	complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:		

CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC

CC Sequence 2201 AA;

Query Match 96.3%; Score 1530; DB 5; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFAHGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDWAHAG 60  
 DB 95 AGITKVPYFAHGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDWAHAG 154  
 QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 120  
 DB 155 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 214  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFILATCVNGVCTVYH 180  
 DB 215 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFILATCVNGVCTVYH 274  
 QY 181 GAGSKTLAAGPKPITQMTYTNVDOLVGMQAPPARSMTCTCGSSDLVLTTRHADVTPVR 240  
 DB 275 GAGSKTLAAGPKPITQMTYTNVDOLVGMQAPPARSMTCTCGSSDLVLTTRHADVTPVR 334  
 QY 241 RRGDSRGSLLSPRPVSYLKSGSGGPLCPGSHAVGIFRAAVCTRGVAKAVDPVPSMET 300  
 DB 335 RRGDSRGSLLSPRPVSYLKSGSGGPLCPGSHAVGIFRAAVCTRGVAKAVDPVPSMET 394  
 QY 301 TMR 303  
 DB 395 TMR 397

RESULT 13

ABG30591 ID ABG30591 standard; protein; 2201 AA.

AC ABG30591;

DT 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 751

FT Misc-difference 882 /note= "Wild type Ser substituted by Gly"

FT Misc-difference 882 /label= Arg, Lys

PN WO200252015-A2.

PD 04-JUL-2002.

PD 20-DEC-2001; 2001WO-CA001643.

PR 22-DEC-2000; 2000US-0257857P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PI Kukulj G, Pause A;

PI WPI; 2002-575382/61.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for

PT evaluating potential inhibitors of HCV replication.  
 CC Claim 3; Page; 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apck12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

CC Sequence 2201 AA;

Query Match 96.3%; Score 1530; DB 5; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFAHGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDWAHAG 60  
 DB 95 AGITKVPYFAHGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDWAHAG 154  
 QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 120  
 DB 155 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 214  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFILATCVNGVCTVYH 180  
 DB 215 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFILATCVNGVCTVYH 274  
 QY 181 GAGSKTLAAGPKPITQMTYTNVDOLVGMQAPPARSMTCTCGSSDLVLTTRHADVTPVR 240  
 DB 275 GAGSKTLAAGPKPITQMTYTNVDOLVGMQAPPARSMTCTCGSSDLVLTTRHADVTPVR 334  
 QY 241 RRGDSRGSLLSPRPVSYLKSGSGGPLCPGSHAVGIFRAAVCTRGVAKAVDPVPSMET 300  
 DB 335 RRGDSRGSLLSPRPVSYLKSGSGGPLCPGSHAVGIFRAAVCTRGVAKAVDPVPSMET 394  
 QY 301 TMR 303  
 DB 395 TMR 397

RESULT 14

ABG30600 ID ABG30600 standard; protein; 2201 AA.

AC ABG30600;

DT 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 882 /label= Arg, Lys

FT Misc-difference 1357  
 FT /note= "Wild type Pro substituted by Leu"  
 FN WO200252015-A2.  
 XX  
 XX 04-JUL-2002.  
 PD  
 XX 20-DEC-2001; 2001WO-CA001843.  
 XX  
 XX 22-DEC-2000; 2000US-0257857P.  
 PR  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA  
 XX Kukulj G, Pause A;  
 XX  
 XX WPI; 2002-575382/61.  
 DR  
 XX  
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 FT  
 XX  
 XX Claim 3; Page: 140pp; English.  
 PS  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations, transduces and/or replicates with greater  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon APOK12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 96.3%; Score 1530; DB 5; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 95 AGITKVPYFVRAHGLIRACMLVRKVAAGHYVQMALMLALTGTYYVDHLTPLODMAHAG 154  
 QY 61 LRLDAVAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 120  
 Db 155 LRLDAVAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 214  
 QY 121 RLAPITTAASQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVYH 180  
 Db 215 RLAPITTAASQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVYH 274  
 QY 181 GAGSKTLAAGPKPITQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 240  
 Db 275 GAGSKTLAAGPKPITQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 334  
 QY 241 RRGSGRSLSPRVSYLKGSSGGPLCPGSHAVGIFRAVCTGTGAKAVDFIPVESMET 300  
 Db 335 RRGSGRSLSPRVSYLKGSSGGPLCPGSHAVGIFRAVCTGTGAKAVDFIPVESMET 394  
 QY 301 TMR 303  
 Db 395 TMR 397

RESULT 15  
 ABG30581  
 ID ABG30581 standard; protein; 2201 AA.  
 XX  
 XX ABG30581;  
 AC  
 XX 21-OCT-2002 (first entry)  
 DT  
 XX  
 XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.  
 DE  
 XX  
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 KM  
 XX Hepatitis C virus.  
 XX  
 XX  
 XX WO200252015-A2.  
 FN  
 XX 04-JUL-2002.  
 PD  
 XX 20-DEC-2001; 2001WO-CA001843.  
 XX  
 XX 22-DEC-2000; 2000US-0257857P.  
 PR  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA  
 XX Kukulj G, Pause A;  
 XX  
 XX WPI; 2002-575382/61.  
 DR  
 XX N-PSDB; ABK88573.  
 DR  
 XX  
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 FT  
 XX  
 XX Disclosure; Page 49-58; 140pp; English.  
 PS  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations, transduces and/or replicates with greater  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon APOK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
 CC  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 96.3%; Score 1530; DB 5; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 95 AGITKVPYFVRAHGLIRACMLVRKVAAGHYVQMALMLALTGTYYVDHLTPLODMAHAG 154  
 QY 61 LRLDAVAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 120  
 Db 155 LRLDAVAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 214  
 QY 121 RLAPITTAASQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVYH 180  
 Db 215 RLAPITTAASQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVYH 274  
 QY 181 GAGSKTLAAGPKPITQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 240  
 Db 275 GAGSKTLAAGPKPITQMTYNNVDOLVGMQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 334

Qy	241	RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET	300
Db	335	RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET	394
Qy	301	TMR	303
Db	395	TMR	397

Search completed: May 6, 2004, 03:30:44  
Job time : 42.2434 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 9.86806 Seconds  
(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-10

Perfect score: 1589  
Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	1532	96.4	3010	1 A45573	genome polypeptide
2	1527	96.1	3010	1 GNVWCU	genome polypeptide
3	1514	95.3	3010	1 GNVWTV	genome polypeptide
4	1486	93.5	3010	1 S18030	genome polypeptide
5	1478	93.0	3010	1 GNVWTC	genome polypeptide
6	1403	88.3	3011	1 S40770	genome polypeptide
7	1398	88.0	3011	1 GNVWVC	genome polypeptide
8	1385	87.2	3011	1 GNVWCH	genome polypeptide
9	1234	77.7	3014	1 UC5620	genome polypeptide
10	1172	73.8	3033	1 Q01303	genome polypeptide
11	1150	72.4	3033	1 GNVW08	genome polypeptide
12	397.5	25.0	3005	2 T08841	genome polypeptide
13	341	21.5	2970	2 T08839	polypeptide - mem
14	101	6.4	600	2 B46642	DNA-directed DNA p
15	99.5	6.3	352	2 G87382	conserved hypochet
16	97.5	6.1	1085	2 T03531	cohn protein homol
17	95.5	6.0	470	2 UC4098	tetracycline 6-hyd
18	93	5.9	660	2 VHWMH2	structural protein
19	92.5	5.8	706	2 S33761	transferrin precu
20	92.5	5.8	716	2 G83612	hypothetical prote
21	91	5.7	904	2 A84212	hypothetical prote
22	90.5	5.7	868	2 H81775	aconitate hydratase
23	90	5.7	2796	2 UC4743	fatty acid synthase
24	89.5	5.6	7463	2 T36248	CDA peptidic synthe
25	88.5	5.5	659	2 B44212	structural protein
26	88	5.5	3414	1 GNVWVE	genome polypeptide
27	87	5.4	3412	1 GNVWVB	genome polypeptide
28	86.5	5.4	470	1 NMIVW8	exo-alpha-sialidase
29	85.5	5.4	348	2 H70549	probable pnhb proc

## ALIGNMENTS

RESULT 1  
A45573  
genome polypeptide - hepatitis C virus (strain UT)

N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001

R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, N

Virus Res. 23, 39-53, 1992

A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: se

A:Reference number: A45573; MUID:92295714; PMID:1318627

A:Accession: A45573

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-3010 <TAN>

A:Cross-references: GB:D11168; GB:D01171; NID:G221612; PID:BA01943.1; PID:G221613

A:Experimental source: HCV-UT

A>Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P:106207)

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polypeptide; serine

F:115/115/Product: capsid protein C #status predicted <CPC>

F:115/191/Product: envelope protein M #status predicted <EM>

F:192-389/Product: major envelope protein B #status predicted <MB>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitisin #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NA>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NB>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 96.4% Score 1532; DB 1; Length 3010;  
Best Local Similarity 95.4% Pred. No. 7.3e-12;  
Matches 289; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAQGLIRACMVRKAGGYYQVAMFKLAALGTIVYDHLTPLODMAHAG 60

DB 904 AAITAMPYFVRAQGLIRACMVRKAGGYYQVAMFKLAALGTIVYDHLTPLODMAHAG 963

QY 61 IEDLVAVEPVFESMEKRIITWGDITACGDIITGLVSAARRGEIILGPADNREGGW 120

DB 964 IEDLVAVEPVFESMEKRIITWGDITACGDIITGLVSAARRGEIILGPADNREGGW 1023

QY 121 RLIAPIITVYSGQTRGLICITISLTGRDNQVEGVEVYVSTATOSFLATCVNGVCMVTFH 180

DB 1024 RLIAPIITVYSGQTRGLICITISLTGRDNQVEGVEVYVSTATOSFLATCVNGVCMVTFH 1083

QY 181 GAGSKTLGPKPIQNTNTNDQDLVGVQAPPGASMTPTCGSSDLYLVTRHADVIYVR 240

DB 181 GAGSKTLGPKPIQNTNTNDQDLVGVQAPPGASMTPTCGSSDLYLVTRHADVIYVR 240

Db 1084 GAGSKTLAAGPKPIITQMTYNVDLVGMHAPPGARSLTFCGSGSDLVLTSHADVIPIVR 1143  
 |||  
 QY 241 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 |||  
 Db 1144 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 |||  
 QY 301 TMR 303  
 |||  
 Db 1204 TMR 1206

## RESULT 2

GNMVTM

genome polypeptide - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001

C:Accession: A39253; PS0086

R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Okoshi, S.; Sugimura, T.; Shimoto

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients

A:Reference number: A39253; MUID:9108550; PMID:2175903

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 &lt;KAT&gt;

A:Cross-references: GB:D0208; NID:G221610; PID:BA44233.1; PID:G221611

R:Kato, N.; Okoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari

A:Reference number: PS0086

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2550-2707 &lt;KAT&gt;

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polypeptide have not been determined.

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin

F:1-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F:116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F:192-389/Product: major envelope protein E #status predicted &lt;NEE&gt;

F:390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F:730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F:1007-1615/Product: hepatitis C virus predicted &lt;NS3&gt;

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted &lt;N4A&gt;

F:1863-2013/Product: nonstructural protein NS4b #status predicted &lt;N4B&gt;

F:2014-3010/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F:156,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 96.1%; Score 1527; DB 1; Length 3010;

Best Local Similarity 94.4%; Pred. No. 2e-123;

Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITVVFVFAAGLIRACMLVRKAGGHVQMAFMKLAALTGYVVDHLTPLODMAHAG 60  
 |||  
 Db 904 AGITVVFVFAAGLIRACMLVRKAGGHVQMAFMKLAALTGYVVDHLTPLODMAHAG 963  
 |||  
 QY 61 LRDLAVALPEVIFSDMEVKIITWGAADTAACDIIISGLPVSARRREIILGPADNFEQGW 120  
 |||  
 Db 964 LRDLAVALPEVIFSDMEVKIITWGAADTAACDIIISGLPVSARRREIILGPADNFEQGW 1023  
 |||  
 QY 121 RLAPITAYSOOTRGLGCIITSLTGDKQVGEVQVSTATOSFLATCNGVCMVYH 180  
 |||  
 Db 1024 RLAPITAYSOOTRGLGCIITSLTGDKQVGEVQVSTATOSFLATCNGVCMVYH 1083  
 |||  
 QY 181 GAGSKTLAAGPKPIITQMTYNVDLVGMHAPPGARSLTFCGSGSDLVLTSHADVIPIVR 240  
 |||  
 Db 1084 GAGSKTLAAGPKPIITQMTYNVDLVGMHAPPGARSLTFCGSGSDLVLTSHADVIPIVR 1143  
 |||  
 QY 241 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Db 1144 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 |||  
 QY 301 TMR 303  
 |||  
 Db 1204 TMR 1206

## RESULT 3

GNMVTM

genome polypeptide - hepatitis C virus (strain Taiwan)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001

C:Accession: A40244

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.

Virology 188, 102-113, 1992

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the

A:Reference number: A40244; MUID:92230206; PMID:1314449

A:Accession: A40244

A:Molecule type: genomic RNA

A:Residues: 1-3010 &lt;CHE&gt;

A:Cross-references: GB:M84754

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural

F:1-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F:116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F:192-389/Product: major envelope protein E #status predicted &lt;NEE&gt;

F:390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F:730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F:1007-1615/Product: hepatitis C virus predicted &lt;NS3&gt;

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted &lt;N4A&gt;

F:1863-2013/Product: nonstructural protein NS4b #status predicted &lt;N4B&gt;

F:2014-3010/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F:156,209,233,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,

Query Match 95.3%; Score 1514; DB 1; Length 3010;

Best Local Similarity 93.1%; Pred. No. 2.6e-122;

Matches 282; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 1 AGITVVFVFAAGLIRACMLVRKAGGHVQMAFMKLAALTGYVVDHLTPLODMAHAG 60  
 |||  
 Db 904 AGITVVFVFAAGLIRACMLVRKAGGHVQMAFMKLAALTGYVVDHLTPLODMAHAG 963  
 |||  
 QY 61 LRDLAVALPEVIFSDMEVKIITWGAADTAACDIIISGLPVSARRREIILGPADNFEQGW 120  
 |||  
 Db 964 LRDLAVALPEVIFSDMEVKIITWGAADTAACDIIISGLPVSARRREIILGPADNFEQGW 1023  
 |||  
 QY 121 RLAPITAYSOOTRGLGCIITSLTGDKQVGEVQVSTATOSFLATCNGVCMVYH 180  
 |||  
 Db 1024 RLAPITAYSOOTRGLGCIITSLTGDKQVGEVQVSTATOSFLATCNGVCMVYH 1083  
 |||  
 QY 181 GAGSKTLAAGPKPIITQMTYNVDLVGMHAPPGARSLTFCGSGSDLVLTSHADVIPIVR 240  
 |||  
 Db 1084 GAGSKTLAAGPKPIITQMTYNVDLVGMHAPPGARSLTFCGSGSDLVLTSHADVIPIVR 1143  
 |||  
 QY 241 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 |||  
 Db 1144 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 |||  
 QY 301 TMR 303  
 |||  
 Db 1204 TMR 1206

## RESULT 4

GNMVTM

genome polypeptide - hepatitis C virus (isolate JKI)

N:Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
A/Variety: isolate JKI  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
C/Accession: S18030; S33570; A48332; S18029  
R/Honda, M.; Kaneko, S.; Masahashi, U.; Kobayashi, K.; Murakami, S.  
Submitted to the EMBL Data Library, September 1991  
A/Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie  
A/Reference number: S18028  
A/Accession: S18030  
A/Molecule type: genomic RNA  
A/Residues: 1-3010 <HON>  
A/Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479  
A/Experimental source: isolate JKI from an individual  
R/Honda, M.; Kaneko, S.; Uenura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
A/Reference number: A48332; MUID:33119270; PMID:8380322  
A/Accession: S33570  
A/Molecule type: genomic RNA  
A/Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DU', 655-761, 'T', 763-782 <HON>  
A/Cross-references: EMBL:X61591  
A/Note: this sequence is inconsistent with the nucleotide translation  
A/Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320  
as Trp, and TTC for residue 771 as Ser  
A/Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EMP>  
F/192-389/Product: major envelope protein E #status predicted <MEB>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepatitis B #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif A (P-loop)  
F/1316-1319/Region: DEXH motif  
F/1616-1662/Product: nonstructural protein NS4 #status predicted <NS4>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,224,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (AS  
Query Match 93.5%; Score 1486; DB 1; Length 3010;  
Best Local Similarity 92.4%; Pred. No. 7e-120;  
Matches 280; Conservative 9; Mismatches 14; Indels 0; Gaps 0;  
QY 1 AGITVYFVFAAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAAG 60  
DB 904 AGITVYFVFAAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAAG 963  
QY 61 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVASARGREILILGPADNFEQGW 120  
DB 964 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVASARGREILILGPADNFEQGW 1023  
QY 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVVSATOSFLATCVNGVCTVPH 180  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVVSATOSFLATCVNGVCTVPH 1083  
QY 181 GAGSKTLGAGKPIITQMTYTNVDQDLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIVR 240  
DB 1084 GAGSKTLGAGKPIITQMTYTNVDQDLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIVR 1143  
QY 241 RRGDSRGLSLPRPVSYLKSSGGPLCPGSHAVGIFFAAVCTRGVAKAVDFIPVESMET 300  
DB 1144 RRGDSRGLSLPRPVSYLKSSGGPLCPGSHAVGIFFAAVCTRGVAKAVDFIPVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

GNMWTG  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
A/Accession: A38465  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
R/Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J  
J. Virol. 65, 1105-1113, 1991  
A/Title: Structure and organization of the hepatitis C virus genome isolated from human  
A/Reference number: A38465; MUID:91140698; PMID:1847440  
A/Accession: A38465  
A/Molecule type: genomic RNA  
A/Residues: 1-3010 <TMK>  
A/Cross-references: EMBL:M58335; NID:G329770; PIDN:AA472945.1; PID:G329771  
A/Experimental source: isolate JKI from an individual  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EMP>  
F/192-389/Product: major envelope protein E #status predicted <MEB>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepatitis B #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif A (P-loop)  
F/1316-1319/Region: DEXH motif  
F/1616-1662/Product: nonstructural protein NS4 #status predicted <NS4>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,224,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224  
Query Match 93.0%; Score 1478; DB 1; Length 3010;  
Best Local Similarity 92.1%; Pred. No. 3.5e-113;  
Matches 279; Conservative 11; Mismatches 13; Indels 0; Gaps 0;  
QY 1 AGITVYFVFAAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAAG 60  
DB 904 AGITVYFVFAAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAAG 963  
QY 61 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVASARGREILILGPADNFEQGW 120  
DB 964 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVASARGREILILGPADNFEQGW 1023  
QY 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVVSATOSFLATCVNGVCTVPH 180  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVVSATOSFLATCVNGVCTVPH 1083  
QY 181 GAGSKTLGAGKPIITQMTYTNVDQDLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIVR 240  
DB 1084 GAGSKTLGAGKPIITQMTYTNVDQDLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIVR 1143  
QY 241 RRGDSRGLSLPRPVSYLKSSGGPLCPGSHAVGIFFAAVCTRGVAKAVDFIPVESMET 300  
DB 1144 RRGDSRGLSLPRPVSYLKSSGGPLCPGSHAVGIFFAAVCTRGVAKAVDFIPVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 6  
S40770  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
A/Accession: S40770; FCI285  
R/Okamoto, H.  
Submitted to the EMBL Data Library, March 1992  
A/Reference number: S40770  
A/Accession: S40770  
A/Molecule type: genomic RNA



A/Residues: 1-3011 <OK>  
 A/Cross-references: EMBL:D10749; NID:g221586; PIDN:BA01582.1; PID:g221587  
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990  
 A/Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A/Reference number: F01284; MUID:91013116; PMID:2170112  
 A/Accession: F01285  
 A/Molecule type: genomic RNA  
 A/Residues: 1-513 <OK>  
 A/Cross-references: GB:D00831; NID:g221511; PIDN:BA00705.1; PID:g221512  
 A/Experimental source: isolate HC-J1  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
 F/2-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <ME>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1319/Region: DEH motif  
 F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 Query Match 88.3%; Score 1403; DB 1; Length 3011;  
 Best Local Similarity 84.8%; Pred. No. 1.1e-112; Indels 0; Gaps 0;  
 Matches 257; Conservative 23; Mismatches 23;  
 Db 1 AGITVFYFAAGLIRACMLVRKAAGHYVQAFMKALNTGYVYDHLTPLDWAHAG 60  
 904 ASLKVYFVFAVQGLIRFCLARKMGHYVQVITLGLNTGYVYVNHLP.LRDWAHNG 963  
 Qy 61 LRLDAVAEVPYISDMVEKITTWADTAAACDITISGIPVARSRRREILLGPADNFEQGM 120  
 Db 964 LRLDAVAEVPVVSOMETKLTWAGADTAAACDITINGIPVARSRRREILLGPADNFEQGM 1023  
 Qy 121 RLAPITAYSQOTRGILGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 180  
 Db 1024 RLAPITAYAQOTRGILGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 1083  
 Qy 181 GAGSKTLAAGKPGPITOMYTNVDOLVGMQAPPGARSMTPTCGSSDLYLTRHADVI 240  
 Db 1084 GAGRTTASPRGPIYQWYTNVDOLVGMQAPPGARSMTPTCGSSDLYLTRHADVI 1143  
 Qy 241 RRGSRGSLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 300  
 Db 1144 RRGSRGSLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 1203  
 Qy 301 TMR 303  
 Db 1204 TMR 1206  
 RESULT 7  
 GNMVOC  
 genome polyprotein - hepatitis C virus (strain HCV-1)  
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 19-Jan-2001  
 C/Accession: A39166; F00403; F00404  
 R/Ochoo, Q.L.; Richman, R.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Col  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
 A/Title: Genetic organization and diversity of the hepatitis C virus.  
 A/Reference number: A39166; MUID:91172826; PMID:1848704  
 A/Accession: A39166  
 A/Molecule type: mRNA  
 A/Residues: 1-3011 <CHO>  
 A/Cross-references: GB:W62321; NID:g329873; PIDN:AAA45676.1; PID:g329874  
 R/Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Feutheuer, J.F.; Follett, E.; Yap, P.L  
 J. Gen. Virol. 73, 1131-1141, 1992

A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
 A/Reference number: F00393; MUID:92268871; PMID:11316939  
 A/Accession: F00403  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CHA>  
 A/Cross-references: DBJ:D10128  
 A/Experimental source: isolates E-b16  
 A/Accession: F00404  
 A/Status: Preliminary  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CH2>  
 A/Experimental source: isolates E-b17  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F/1-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <ME>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1319/Region: DEH motif  
 F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,224  
 Query Match 88.0%; Score 1398; DB 1; Length 3011;  
 Best Local Similarity 84.5%; Pred. No. 2.9e-112; Indels 0; Gaps 0;  
 Matches 256; Conservative 24; Mismatches 23;  
 Db 1 AGITVFYFAAGLIRACMLVRKAAGHYVQAFMKALNTGYVYDHLTPLDWAHAG 60  
 904 ASLKVYFVFAVQGLIRFCLARKMGHYVQVITLGLNTGYVYVNHLP.LRDWAHNG 963  
 Qy 61 LRLDAVAEVPYISDMVEKITTWADTAAACDITISGIPVARSRRREILLGPADNFEQGM 120  
 Db 964 LRLDAVAEVPVVSOMETKLTWAGADTAAACDITINGIPVARSRRREILLGPADNFEQGM 1023  
 Qy 121 RLAPITAYSQOTRGILGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 180  
 Db 1024 RLAPITAYAQOTRGILGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 1083  
 Qy 181 GAGSKTLAAGKPGPITOMYTNVDOLVGMQAPPGARSMTPTCGSSDLYLTRHADVI 240  
 Db 1084 GAGRTTASPRGPIYQWYTNVDOLVGMQAPPGARSMTPTCGSSDLYLTRHADVI 1143  
 Qy 241 RRGSRGSLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 300  
 Db 1144 RRGSRGSLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 1203  
 Qy 301 TMR 303  
 Db 1204 TMR 1206  
 RESULT 8  
 GNMVCH  
 genome polyprotein - hepatitis C virus (strain H)  
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 A/Note: host Homo sapiens (man)  
 C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C/Accession: A36814; A41546  
 R/Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 submitted to GenBank, July 1992  
 A/Description: Genomic structure of the human prototype strain H of hepatitis C virus: cc  
 A/Reference number: A36814  
 A/Accession: A36814  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3011 <INC>

A;Cross-references: GB:M67463; NID:9329737; PIDN:AAA5534.1; PID:9329738  
 R;Inchouse, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison  
 A;Reference number: A41546; MUID:92052256; PMID:1558900  
 A;Contents: annotation  
 A;Note: neither amino acid nor nucleotide sequence is given  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydroxylase; nonstructural  
 F;1115/Product: capsid protein C #status predicted <CPC>  
 F;116-191/Product: envelope protein M #status predicted <EPM>  
 F;192-189/Product: major envelope protein E #status predicted <MEP>  
 F;380-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F;1230-1237/Product: nucleotide-binding motif A (P-loop)  
 F;1312-1317/Region: nucleotide-binding motif B  
 F;1316-1319/Region: DEXH motif  
 F;1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F;196-209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 87.2%; Score 1385; DB 1; Length 3011;  
 Best Local Similarity 83.8%; Pred. No. 3,9e-111;  
 Matches 254; Conservative 26; Mismatches 23; Indels 0; Gaps 0;

QY 1 AGTTVPYPAAGLIRACMLVRKAGGHVYQAFMKALALGTGYVDHLTPQDMAHAG 60  
 DB 904 ASLKKYFVFRVQGLALICARLAKAGHYQMAIIKGLTICVYNHAPLRDAHNG 963  
 QY 61 LRDIAVVEPVISDMVKIITWGADTPAACGDIISGLPVASARGRETLGPDNFGQGM 120  
 DB 964 LRDIAVVEPVISDMVKIITWGADTPAACGDIISGLPVASARGRETLGPDNFGQGM 1023  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSPFATCNGVCTVFFH 180  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSPFATCNGVCTVFFH 1083  
 QY 181 GAGSKITAGKRPITQWYTNVDQILNGWQAPPGARSMPTCTCGSSDLVYTRHADVIPRR 240  
 DB 1084 GAGSKITAGKRPITQWYTNVDQILNGWQAPPGARSMPTCTCGSSDLVYTRHADVIPRR 1143  
 QY 241 RRGDSRGLSPRPVSYLKSSGGPILCPGSHANGIRAAVCTRGVAKADVFPVEMET 300  
 DB 1144 RRGDSRGLSPRPVSYLKSSGGPILCPGSHANGIRAAVCTRGVAKADVFPVEMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 9

JC5620 genome polyprotein - hepatitis C virus (isolate EUH1460)  
 N;Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C;Accession: J05620  
 R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
 A;Reference number: JC5620; MUID:9736593; PMID:9223433  
 A;Accession: J05620  
 A;Molecule type: mRNA  
 A;Residues: 1-3014 <CHA>  
 A;Cross-references: GB:Y13314  
 A;Experimental source: genotype 5a, which predominates in South Africa  
 A;Note: the translation of the nucleotide sequence is not complete in this paper  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydroxylase; nucleotide binding; P-loop; polyprotein; serin  
 F;2-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product: envelope protein M #status predicted <EPM>  
 F;192-189/Product: major envelope protein E #status predicted <MEP>  
 F;384-408/Region: hypervariable #status predicted  
 F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1008-1616/Product: hepatitisin #status predicted <NS3>  
 F;1231-1338/Region: nucleotide-binding motif A (P-loop)  
 F;1313-1318/Region: nucleotide-binding motif B  
 F;1311-1320/Region: DEXH motif  
 F;1611-1863/Product: nonstructural protein NS4 #status predicted <NS4>  
 F;1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F;2015-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F;2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 77.7%; Score 1234; DB 1; Length 3014;  
 Best Local Similarity 72.8%; Pred. No. 4.7e-98;  
 Matches 219; Conservative 43; Mismatches 39; Indels 0; Gaps 0;

QY 3 IRKVPYPAAGLIRACMLVRKAGGHVYQAFMKALALGTGYVDHLTPQDMAHAG 62  
 DB 907 LKVPYFRLARLLMLCLAKLTVGKXYQALALHGLTGYIDHLAPKDMASGLR 966  
 QY 63 DLAVAVEPVISDMVKIITWGADTPAACGDIISGLPVASARGRETLGPDNFGQGM 122  
 DB 967 ELTVATEPVISDMVKIITWGADTPAACGDIISGLPVASARGRETLGPDNFGQGM 1026  
 QY 123 LAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSPFATCNGVCTVFFH 182  
 DB 1027 LAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSPFATCNGVCTVFFH 1086  
 QY 183 GSKITAGKRPITQWYTNVDQILNGWQAPPGARSMPTCTCGSSDLVYTRHADVIPRR 242  
 DB 1087 GSKITAGKRPITQWYTNVDQILNGWQAPPGARSMPTCTCGSSDLVYTRHADVIPRR 1146  
 QY 243 RRGDSRGLSPRPVSYLKSSGGPILCPGSHANGIRAAVCTRGVAKADVFPVEMET 302  
 DB 1147 RRGDSRGLSPRPVSYLKSSGGPILCPGSHANGIRAAVCTRGVAKADVFPVEMET 1206  
 QY 303 R 303  
 DB 1207 R 1207

RESULT 10

JQ1303 genome polyprotein - hepatitis C virus (isolate HC-26)  
 N;Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C;Accession: JQ1303  
 R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurali, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;  
 J. Gen. Virol. 72, 2697-2704, 1991  
 A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a huma  
 A;Reference number: JQ1303; MUID:92044440; PMID:1658196  
 A;Accession: JQ1303  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3033 <OKA>  
 A;Cross-references: GB:U00944; NID:9221650; PIDN:BA00792.1; PID:9221651  
 A;Experimental source: isolate HC-26 from a Japanese individual  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydroxylase; P-loop; polyprotein; serine proteinase; transme  
 F;2-115/Product: capsid protein C #status predicted <CPC>  
 F;116-191/Product: envelope protein M #status predicted <EPM>  
 F;192-189/Product: major envelope protein E #status predicted <MEP>  
 F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1011-1619/Product: hepatitisin #status predicted <NS3>  
 F;1316-1321/Region: nucleotide-binding motif B  
 F;1320-1323/Region: DEXH motif  
 F;1620-1866/Product: nonstructural protein NS4 #status predicted <NS4>  
 F;1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>



QY 272 HAAGFRAAVCTRGV-----AKAVDFIPVES 297  
 Db 1124 HAVGGM-VSVLHRGVKTKYRVKPMETLPRDS 1155

## RESULT 13

T08839  
 polyprotein - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: T08839  
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <ERR>  
 A/Cross-references: EMBL:AF023424; NID:G2828597; PIDN:AC40501.1; PID:G2828598  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 21.5%; Score 341; DB 2; Length 2970;  
 Best Local Similarity 30.5%; Pred. No. 1e-20;  
 Matches 85; Conservative 46; Mismatches 96; Indels 52; Gaps 9;

QY 56 MAHAG-----LRDLAAVEPVIFSDMEVKIITGADPTAACGDIISGLPVARSRGRE 106  
 Db 891 YAHAGVTRTAEELRQMGFALEPVAVHPEDCAMVRAATLSGGGSHGKPVAAKRGDE 950  
 QY 107 ILIGPADNFEQGQWRL-----LAPITAYSQQTRGLGCIITSLTGRDKNQVGEVQVVS 160  
 Db 951 VLIGVLNGV---WELPFGVPTAPVYVH-HHGKGFQGVKTSMTGMDTEHGVVVLG 1005  
 QY 161 TATQSFATCVNGCVGVEFHGAGSKTLAGKRPITQMYTNVDDIVGQWAPPARASMTPC 220  
 Db 1006 TSTIRSGTCVNGVMTYTHGSSNKTLLAQMGPVNSRWASDVAVYPLPVAKKCEPC 1065  
 QY 221 TCGSSDLYLVTRHADVIPRRRQDSRSGSLSS-----PRPVSYLKSSSGSPILCP 269  
 Db 1066 KCGQGVWVI-----RND--GALCHGTIGRTVELDLPALCDPFGSSGSPILCD 1112  
 QY 270 SGHAGVGFRAAVCTRG-----YAKAVDFIPVESMTT 301  
 Db 1113 EGHAVGML-ISVLRGSRVGTGIRYTKPMETLPRALTHT 1150

## RESULT 14

B46642  
 DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68k chain  
 C/Species: Mus musculus (house mouse)  
 C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
 C/Accession: B46642  
 R/Miyazawa, H.; Ikumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.  
 J. Biol. Chem. 268, 8111-8122, 1993  
 A/Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase alpha  
 A/Reference number: A46642; MUID:93216786; PMID:8463324  
 A/Accession: B46642  
 A/Status: preliminary  
 A/Molecule type: mRNA; protein  
 A/Residues: 1-600 <MTY>  
 A/Cross-references: GB:D13546; NID:9303658; PIDN:BA02746.1; PID:9303659  
 A/Experimental source: EM3A cells  
 A/Note: Sequence extracted from NCBI backbone (NCBI:129148, NCBI:P:129149)  
 C/Keywords: nucleotidyltransferase

Query Match 6.4%; Score 101; DB 2; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 0.85;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 15 LIRACMLVRAAGHYQM-AFMKALIT-----GTYYVDH-----TPIQDNA 57

Db 27 LAELCVLYRQEDGAVSELIARCTAGKTCITVDILNSFEVEVLNKLKSKAMHASKSCSG 86  
 QY 58 HAGLADLAVAPEVIFSDMEVKIITWGDPTAACGDI--ISGLP-----VSARGREI 107  
 Db 87 HAGTRDI-VSIQELLEAESEETLLSSYTPPSKGPLKRVSTPETPLTKSVAAASPRO- 144  
 QY 108 LIGPADNFEQGQWRLAPITAYSQQTRGLGCIITSLTGRDKNQVGEVQVVSSTATQSF 167  
 Db 145 LSPFS-----FSPBATPSCK-----YTSRNR-----GEVITTFGSAQ--- 178  
 QY 168 ATCVNGCVTFHAGSKTL--AGKRPITQMYTNVDDIVG 207  
 Db 179 -----GLSWSGRGSGSVSLKVVGDPEPLTGSYKAMFOQLMG 215

## RESULT 15

G87392  
 conserved hypothetical protein CCL155 (imported) - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: G87392  
 R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: G87392  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-353 <STO>  
 A/Cross-references: GB:AE005673; NID:913422473; PIDN:AAK23139.1; GSPDB:GN00148  
 C/Genetics:  
 A/Genes: CCL155

Query Match 6.3%; Score 99.5; DB 2; Length 353;  
 Best Local Similarity 21.8%; Pred. No. 0.59;  
 Matches 72; Conservative 36; Mismatches 114; Indels 109; Gaps 13;

QY 7 PYFVRAQGLIRACMLVRKAA-----GGHYV-----QMAFMKALITGTYYVDH 50  
 Db 65 PLAVLAGGFARFQSLARBSAIVMARASGSGYRIYGVAVPAVAVMLDALCGVLAIRA 124  
 QY 51 TP-LQDMAHAGRLDAVAPEVIFSDMEVKIITWGDPTAACGDIISGLPVARSRGREIL 109  
 Db 125 DEPLADW-WRNTTPVAERKEPVPRTFRAGADLVIGANASADRTITGVTIFRRDSKGLIV 183  
 QY 110 -----GPADNFEQGQWRLAPITAYSQQTRGLGCIITSLTGRDKNQVGEVQVVSSTATQS 165  
 Db 184 EKVYEAARARYDQKATLTLPKT-----TRFADLSQAATA-- 219  
 QY 166 FLATCVNGCVTFHAGSKTLTAGKRPITQMYTNVDDIVGQWAPPARASMTPTCCSS 225  
 Db 220 -----ATSWP-----TALRPDVOVGLFGDDSDMPAAS----- 246  
 QY 226 DLYVTRHADVIPRRRQDSRSGSLSPRVSY-----LKGSSSGP-----LLCPSGHVG 275  
 Db 247 -----ARRALENG--SDRPESRYATHLQAFASPSVSLVMLLSAPVALA 290

Search completed: May 6, 2004, 09:37:15  
 Job time: 10.8681 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.32568 Seconds  
(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-10  
Sequence: 1 AGITKPYFVRAGLIRACV.....RGVAKAVDPFVPSMETTMR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1532	96.4	3010 1	POLG_HCVJT
2	1527	96.1	3010 1	POLG_HCVJA
3	1514	95.3	3010 1	POLG_HCVTW
4	1478	93.0	3010 1	POLG_HCVBK
5	1398	88.0	3011 1	POLG_HCVL
6	1385	87.2	3011 1	POLG_HCVH
7	1172	73.8	3033 1	POLG_HCVJ6
8	1150	72.4	3033 1	POLG_HCVJ8
9	101	6.4	600 1	DP02_MOUSE
10	93	5.9	660 1	VST2_HEVBU
11	93	5.9	660 1	VST2_HEVBU
12	92.5	5.8	706 1	TRFE_HORSE
13	89.5	5.6	659 1	VST2_HEVME
14	88	5.5	3414 1	POLG_TBSEV
15	87	5.5	3414 1	POLG_TBSEV
16	87	5.5	3414 1	POLG_TBSEV
17	86.5	5.4	470 1	NRAM_IAMHM
18	86	5.4	434 1	TOLE_CHILTE
19	85	5.3	470 1	NRAM_IATRA
20	85	5.3	470 1	HELS_METMA
21	84.5	5.3	347 1	MDHM_EUCGU
22	84.5	5.3	1705 1	PTPV_MOUSE
23	84	5.3	309 1	UCP2_RAT
24	84	5.3	339 1	GPDA_COREF
25	83.5	5.3	538 1	NRAM_IARUE
26	83.5	5.3	538 1	DAC_ACTSP
27	83.5	5.3	854 1	PMW2_SCHPO
28	83	5.2	485 1	VST2_HEVRY
29	83	5.2	660 1	VST2_HEVRY
30	82.5	5.2	453 1	NRAM_IAMWL
31	82	5.2	309 1	UCP2_MOUSE
32	82	5.2	403 1	PGK_CHLMU
33	82	5.2	612 1	AMYG_ASPOR

34	81.5	5.1	398 1	TRNU_AGRIS	Q8u9ms agrobacteri
35	81.5	5.1	1022 1	CA26_CHICK	P15988 gallus gall
36	81	5.1	209 1	PAAD_PSEAE	O9ux08 pseudomonas
37	81	5.1	350 1	PE24_ARATH	O9zvt04 arabidopsis
38	81	5.1	470 1	NRAM_IAMUS	P03469 influenza a
39	81	5.1	730 1	HELS_METAC	O8t139 methanosarc
40	80	5.0	309 1	UCP2_HUMAN	P55851 homo sapien
41	80	5.0	326 1	PANE_RHIL0	Q987n5 rhizobium 1
42	80	5.0	961 1	ATCU_YERPE	O8zcat7 yersinia pe
43	80	5.0	3491 1	ERY1_SACER	O01331 saccharopol
44	79.5	5.0	339 1	CRTB_RHCCA	P17056 rhodobacter
45	79.5	5.0	453 1	GAG_AVTMD	P06444 avian myelo

ALIGNMENTS

RESULT 1  
POLG\_HCVJT STANDARD; PRT; 3010 AA.  
AC 000269;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate HC-JT) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31642;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92295714; PubMed=1318627;  
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
RT Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
RT "Molecular cloning of hepatitis C virus genome from a single Japanese  
RT carrier: sequence variation within the same individual and among  
RT infected individuals.";  
RL Virus Res. 23:39-53(1992).  
CC -!- FUNCTION: The small proteins NS2, NS2B, NS4A and NS4B are  
CC hydrophobic, suggesting a possible membrane-related function. NS3  
CC and NS5 may play a role in the viral RNA replication.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA) (N).  
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
CC lipoprotein envelope. The envelope consists of two proteins:  
CC protein M and glycoprotein E. The nucleocapsid is a complex of  
CC protein N and RNA.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>).  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
DR EMBL; D11168; BA01943.1; -.  
DR PIR; A45573; A45573.  
DR MEROPS; S29.001; -.  
DR MEROPS; U39.001; -.  
DR InterPro; IPR009003; Cys\_ser\_trypsin.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV\_capsid.

DR	InterPro: IPR002521; HCV_core.
DR	InterPro: IPR002519; HCV env.
DR	InterPro: IPR002531; HCV_NS1.
DR	InterPro: IPR002518; HCV_NS2.
DR	InterPro: IPR000745; HCV_NS4.
DR	InterPro: IPR001490; HCV_NS4b.
DR	InterPro: IPR002868; HCV_NS5A.
DR	InterPro: IPR002166; HCV_NS4D.
DR	InterPro: IPR001650; Helicase_C.
DR	InterPro: IPR004109; Peptidase_C9.
DR	InterPro: IPR007095; RNA_pol_d5_PS.
DR	InterPro: IPR007094; RNA_pol_NSvtr.
DR	Pfam: PF01543; HCV_core; 1.
DR	Pfam: PF01542; HCV_core; 1.
DR	Pfam: PF01539; HCV env; 1.
DR	Pfam: PF01560; HCV_NS1; 1.
DR	Pfam: PF02937; HCV_NS2; 1.
DR	Pfam: PF02907; HCV_NS3; 1.
DR	Pfam: PF01006; HCV_NS4a; 1.
DR	Pfam: PF01001; HCV_NS4b; 1.
DR	Pfam: PF01506; HCV_NS5a; 1.
DR	Pfam: PF00271; helicase_C; 1.
DR	Pfam: PF00998; Viral RdRp; 1.
DR	ProDom: PD186062; HCV_NS1; 1.
DR	SMART; SM00487; DEXDC; 1.
KW	Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW	Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW	Transmembrane; Nonstructural
FT	INIT_MET 1 1
FT	CHAIN 1 115
FT	CHAIN 116 191
FT	CHAIN 192 383
FT	CHAIN 384 729
FT	CHAIN 730 1006
FT	CHAIN 1007 1615
FT	CHAIN 1616 1682
FT	CHAIN 1683 2013
FT	CHAIN 2014 3010
FT	TRANSMEM 347 369
FT	ACT_SITE 1083 1083
FT	ACT_SITE 1107 1107
FT	ACT_SITE 1165 1165
FT	NP_BIND 1230 1237
FT	SITE 1316 1319
FT	CARBOHYD 196 196
FT	CARBOHYD 209 209
FT	CARBOHYD 234 234
FT	CARBOHYD 250 250
FT	CARBOHYD 305 305
FT	CARBOHYD 417 417
FT	CARBOHYD 423 423
FT	CARBOHYD 430 430
FT	CARBOHYD 448 448
FT	CARBOHYD 532 532
FT	CARBOHYD 540 540
FT	CARBOHYD 556 556
FT	CARBOHYD 576 576
FT	CARBOHYD 623 623
FT	CARBOHYD 645 645
FT	CARBOHYD 2041 2041
FT	CARBOHYD 2077 2077
FT	CARBOHYD 2240 2240
FT	CARBOHYD 2529 2529
FT	CARBOHYD 2768 2768
SEQUENCE	3010 AA; 326573 MW; 9A1C77435D642BB CRC64;

```

Query Match          96.4%; Score 1532; DB 1; Length 3010;
Best Local Similarity 95.4%; Pred..NO. 1.6e-124;
Matches 289; Conservative 6; Mismatches 8; Indels 0; Gaps 0
QY      1 AGITKVPYFVRAGLIRACMLVKAAAGHYQVAMFKIALITGYVTDHILTPQDPANHG 60

```

```

Db      904 AATTAAEYFPRAGSLIRACMLVAKVGAVGHVQVAEMKLAALTGTYYADLTPQLDMAHAG 963
Oy      61 LRLIAVAEPVITSNDEWEVKITTMGADLTAAAGDIIISGLPSSARGRREILIGPANFEQGGM 120
Db      964 LRLIAVAEPVITSNDEWEVKITTMGADLTAAAGDIIISGLPSSARGRREILIGPANFEQGGM 1023
Oy      121 RLAPIPAYSQQRRLGLLCITTSITSGDKNQVEGEVQVVSTATOSFLATCVNGVCWTFEH 180
Db      1024 RLAPIPITAYQQIRGLLGCLVTSLTGGDNKNQVEGEVQVVSTATQSFLATCVNGVCWTFEH 1083
Oy      181 GAGSKTLAGEPKPTQMTYNVDODLVGMQAPPGGARSMTPCTCGSSDLXYLTRADYTPVR 240
Db      1084 GAGSKTLAGEPKPTQMTYNVDODLVGMQAPPGGARSLTPTCGSSDLXYLTRADYTPVR 1143
Oy      241 RRGGSRSLSLSPRVSYLKSSGGGPLICPSGHAVGIFRAVCTRGVAKADVFIPIVESMET 300
Db      1144 RRGGRSLSLSPRVSYLKSSGGGPLICPSGHAVGIFRAVCTRGVAKADVFIPIVESMET 1203
Oy      301 TMR 303
Db      1204 TMR 1206

RESULT 2
POLG_HCVJA STANDARD; PRT; 3010 AA.
ID POLG_HCVJA          STAN     PR    3010
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein (contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
DE (BC 3.4.22.-.); Protease/helicase NS3 (P70) (Hepadnavirin)
DE (EC 3.4.21.96); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OC Hepatitis C virus (Isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91089550; PubMed=2175503;
RA Kato N., Hijiata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis."
RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijiata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome."
RT FEBS Lett. 280:325-328(1991).
RN [3]
RP FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
RP hydrophobic, suggesting a possible membrane-related function. NS3
RP and NS5 may play a role in the viral RNA replication.
RP -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
RP precursor polypeptide, commonly with Asp or Glu in the P6
RP position, Cys or Thr in P1 and Ser or Ala in P1'.
RP -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
RP {RNA} (N).
RP -I- SUBUNIT: The viron of this virus is a nucleocapsid covered by a
RP lipoprotein envelope. The envelope consists of two proteins:
RP protein M and glycoprotein E. The nucleocapsid is a complex of
RP protein C and mRNA.
RP -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its

```



[illegible]

FT	CAROHXD	2077	2077	N-LINKED (GLCNAC...)	(POTENTIAL)
FT	CAROHXD	2240	2240	N-LINKED (GLCNAC...)	(POTENTIAL)
FT	CAROHXD	2768	2768	N-LINKED (GLCNAC...)	(POTENTIAL)
SQ	SEQUENCE	3010 AA;	327017 MM;	AA933794F46DB185	CRG64;
 Query Match					
QY			96.1%;	Score 1527;	DB 1; Length 3010;
Db			Best Local Similarity	94.4%;	Pred. No. 4,4e-124;
	Matches	286;	Conservative	11;	Mismatches 6; Indels 0; Gaps 0;
QY	1	AGITVPPFPVBOGIIRACLVKRAAGHYVQMAFMKLALTGTVYDHLTPLODMAHAG	60		
Db	904	AGITREVPFYFAOQGIIRACMKVRKAGGHYVQMAFMKLALTGTVYDHLTPLRDMAHAG	963		
QY	61	LRLDAVAEPEVFPSMEKKIITWAGDTTACGDIIISGLPVSRARRGEILLGPADSGEGGW	120		
Db	964	LRLDAVAEPEVFPSMETKYLITWGADTTACDIIISGLPVSRARGEILLGPADSGEGGW	1023		
QY	121	RLAAPTATYSOOTRGLGCIIITSLTGRDNQVEGVGYVSTATOSFLATCAVCWCMTYFH	180		
Db	1024	RLAPAITATYSOOTRGLGCIIITSLTGRDNQVDDEVQVLSTAIATGSFLATCAVCWCMTYTH	1083		
QY	181	GAGSHTLAGEPGPIPTQMYNTNDOLVGQAEPGARSMPTCTCGSDLYLTRHADVIIVR	240		
Db	1084	GAGSHTLAGEPGPIPTQMYNTNDOLVGPAEPGARSMPTCTCGSDLYLTRHADVIIVR	1143		
QY	241	RRCDSRGSLSPRPVSPLYKSSGGPILCPESHAVGIFPAAVCTSGVAKAVDFIVESMET	300		
Db	1144	RRCDSRGSLSPRIISTYKSSGGPILCPSHHVIGIFRAVCTGTGAKAVDFFIVESMET	1203		
QY	301	TMR 303			
Db	1204	TMR 1206			
 RESULT 3					
ID	POLG_HCVTM	STANDARD;	PRT;	3010 AA.	
AC	P29846;				
DT	01-APR-1993 (Rel. 25, Created)				
DT	01-APR-1993 (Rel. 25, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Genome polypeptide [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)]				
DE	(EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].				
DE	Hepatitis C virus [Isolate taiwan] (HCV).				
OS	Virusess; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.				
OC	NCBI_TaxID=31645;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92230205; PubMed=1314449;				
RA	Chen P.C., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;				
RT	"The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' terminus of viral genomic and antigenomic RNA.";				
RL	Virology 186:102-113(1992).				
CC	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.				
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.				
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).				
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.				
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.				

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M84754; -, NOT\_ANNOTATED\_CDS.  
DR PIR; A40244; GNMVTM.  
DR PDB; 1N63; 25-FEB-03.  
DR PDB; 1N63; 08-APR-98.  
DR MEROPS; S29.001; -.  
DR MEROPS; U39.001; -.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR000745; HCV\_NS4.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS4d.  
DR InterPro; IPR002166; HCV\_RdRp.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; peptidase\_C29.  
DR InterPro; IPR007094; RNA\_pol\_DS\_PS.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS4d; 1.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRp; 1.  
DR Pfam; PF016062; HCV\_NS1; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR KEGG; K04487; Glycoprotein; Transferrin; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;  
KM 3D-structure.  
FT INT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3010  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1237 1237  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 233 233  
FT CARBOHYD 234 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 546 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDEB215 CAC64;

Query Match 95.3%; Score 1514; DB 1; Length 3010;  
Best Local Similarity 93.1%; Pred. No. 5; 9e-123;  
Matches 282; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 1 AGITKVPYFPAQCLIPACMLVRAAGGVVQAEPKALATGYVYDHLTPDQMAHAG 60  
DB 904 AGITRIPYFPAQCLIPACMLVRAAGGVVQAEPKALATGYVYDHLTPDQMAHAG 963  
QY 61 LRDLAVAVEPVESDMVEKIIITWQADTAACDIIISGLPVASARRGREIILGPADNPFSGQW 120  
DB 964 LRDLAVAVEPVESDMVEKIIITWQADTAACDIIISGLPVASARRGREIILGPADNPFSGRW 1023  
QY 121 RLAPITRAYSCQRTGLICITISLTERDKNOVEGVVSTATQSFATCGVCTVPH 180  
DB 1024 RLAPITRAYSCQRTGLICITISLTERDKNOVEGVVSTATQSFATCGVCTVPH 1083  
QY 181 GAGSKTLAAGPKPPTQWYTVNDQPLVGMQAPPGARSMTPCTCGSSDLYLTRADVIVR 240  
DB 1084 GAGSKTLAAGPKPPTQWYTVNDQPLVGMQAPPGARSMTPCTCGSSDLYLTRADVIVR 1143  
QY 241 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIRAAVCTRGVAKANDFIPVSEMT 300  
DB 1144 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIRAAVCTRGVAKANDFIPVSEMT 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 4  
POLG HCVBK STANDARD; PRT; 3010 AA.  
ID POLG HCVBK  
AC P26663;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate BK) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepatitis virus.  
OX NCBI\_Taxid=1105;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91140698; PubMed=1847440;  
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,  
RA Onishi E., Andoh T., Yoshida I., Okayama H.;  
RT "Structure and organization of the hepatitis C virus genome isolated  
RT from human carriers";  
RJ J. Virol. 65:1105-1113 (1991).  
RP SEQUENCE OF 1487-1500.  
RX MEDLINE=9635224; PubMed=8647104;  
RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetcky L.;  
RT "Non-structural protein 3 of hepatitis C virus inhibits





```

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 6578C9447FCBFAF9 CRC64;

Query Match 88.0%; Score 1398; DB 1; Length 3011;
Best Local Similarity 84.5%; Pred. No. 7e-113;
Matches 256; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

QY 1 AGIKVYFVPAQGLIACMIVRAAGHYVQAFMLALITGVYVDHITPLDVAHAG 60
D 904 ASLKVYFVPAQGLIACMIVRAAGHYVQAFMLALITGVYVDHITPLDVAHAG 963
QY 61 LRDLAVAVEPVPFQSMETKLTWAGADTAACDITISGLPVASRRREITLGPANFEGQW 120
D 964 LRDLAVAVEPVPFQSMETKLTWAGADTAACDITISGLPVASRRREITLGPANFEGQW 1023
QY 121 RLAPITAYSQOTGLIGCITTSITGDKNQVEGEVQVSTATQSFATVNGVCTVPH 180
D 1024 RLAPITAYSQOTGLIGCITTSITGDKNQVEGEVQVSTATQSFATVNGVCTVPH 1083
QY 181 GAGSKTLAPGKPIQYVTVNDQVLVGMQAPGASMTPTCCSSDLVLTTRHADVIVR 240
D 1084 GAGSKTLAPGKPIQYVTVNDQVLVGMQAPGASMTPTCCSSDLVLTTRHADVIVR 1143
QY 241 RRGDSRGSLLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAKAVDFIVESMET 300
D 1144 RRGDSRGSLLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAKAVDFIVESMET 1203
QY 301 TMR 303
D 1204 TMR 1206

RESULT 6
POLG HCVA STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (NS3, 4.99.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (NS3, 4.21.38); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE Hepatitis C virus (isolate H) (HCV).
OS Hepatitis C virus positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitis C virus;
OC NCBI_TaxID=11108;
RN [1]
RX MEDLINE=92052256; PubMed=1658800;
RA Inchausti G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.",
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

```

```

RX MEDLINE=9731322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.",
RL Nat. Struct. Biol. 4:463-467(1997).
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RA Kim J.L., Morgensstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT Hepatitis C virus NS3 helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding";
RL Structure 6:89-100(1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND NS5A.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
DR EMBL; M67463; AAA45534.1; -.
DR PIR; A36814; GNVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR TRANSFAC; T04155; -.
DR INTERPRO; IPR009003; Cys_Ser_trypsin.
DR INTERPRO; IPR001410; DEAD.
DR INTERPRO; IPR002522; HCV_capsid.
DR INTERPRO; IPR002521; HCV_core.
DR INTERPRO; IPR002519; HCV_env.
DR INTERPRO; IPR002531; HCV_NS1.
DR INTERPRO; IPR002518; HCV_NS2.
DR INTERPRO; IPR000745; HCV_NS4A.
DR INTERPRO; IPR001490; HCV_NS4B.
DR INTERPRO; IPR002868; HCV_NS5A.
DR INTERPRO; IPR002166; HCV_NS5B.
DR INTERPRO; IPR001650; Helicase_C.
DR INTERPRO; IPR004109; Peptidase_C29.
DR INTERPRO; IPR007095; RNA_pol_DS_PS.
DR INTERPRO; IPR007094; RNA_pol_PSVLr.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.

```



CC Hepacivirus.  
 OX NCBI\_Taxid=11113;  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=92044440; PubMed=1658196;  
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kuzai K., Iizuka H.,  
 RA Machida A., Miyakawa Y., Mayumi M.;  
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 from a human carrier: comparison with reported isolates for conserved  
 RT and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704 (1991).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 hydrophobic, suggesting a possible membrane-related function. NS3  
 and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 precursor polypeptide, commonly with Asp or Glu in the P6  
 position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 lipoprotein envelope. The envelope consists of two proteins:  
 protein M and glycoprotein E. The nucleocapsid is a complex of  
 protein C and RNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D00944; BAA00792.1; -;  
 DR PIR: J01303; J01303.  
 DR HSSP: P27958; 1HE1.  
 DR MEROPS: S29.001; -;  
 DR MEROPS: U39.001; -;  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR InterPro: IPR01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SW00487; DEXDC; 1.  
 KW Polypeptide, Glycoprotein, Transferase, RNA-directed RNA polymerase,  
 KW Core protein, Coat protein, Envelope protein, Helicase, ATP-binding,  
 KW Transmembrane, Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 MATRIX PROTEIN (POTENTIAL).  
 -----

FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1067 1087  
 FT ACT\_SITE 1131 1131  
 FT ACT\_SITE 1169 1169  
 FT ACT\_SITE 1234 1241  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 156 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2811 2811  
 SQ SEQUENCE 3033 AA; 329165 MW; P957F5CIA273BE9E CRC64;  
 Query Match 73.8%; Score 1172; DA 1; Length 3033;  
 Best Local Similarity 69.4%; Pred. No. 3e-93;  
 Matches 209; Conservative 44; Mismatches 48; Indels 0; Gaps 0;  
 QY 3 ITKVEYFPAQGLIRACMLVYKAGGHVYQAFMKLALITGYVDHITPLQDPAHAGLR 62  
 DB 910 ITRVPEYFRAHALIMCTMVRHLRAGRYVQWVLLALGRTGYTYIDHITLPSDMAANGLR 969  
 QY 63 DLAAVVEVIFSDDEKVIITWGADTPAACGDIISGLPVARSREITLIPANFEGQWRL 122  
 DB 970 DLAAVVEVIFSDDEKVIITWGADTPAACGDIISGLPVARSREITLIPANFEGQWRL 1029  
 QY 123 LAPITAYSQOTRGILGICITISLTVGRKNQVEGEVQVAVSTATQSFILATCVNGCVTFHGA 182  
 DB 1030 LAPITAYSQOTRGILGICITISLTVGRKNQVEGEVQVAVSTATQSFILATCVNGCVTFHGA 1089  
 QY 163 GSKTLAGKSPITOMVTVNDDIVGQWAPPGARSMPTCTGSSDPLVYTRADYIPVARR 242  
 DB 1090 GSKTLAGKSPITOMVTVNDDIVGQWAPPGARSMPTCTGSSDPLVYTRADYIPVARR 1149  
 QY 243 GDSRGLSPRPVSYLTKSSGGPPLCPGSAVAGIIFRAAVCTRGVAKAVDIPVSMETTM 302  
 DB 1150 GDSRGLSPRPVSYLTKSSGGPPLCPGSAVAGIIFRAAVCTRGVAKAVDIPVSMETTM 1209  
 QY 303 R 303  
 DB 1210 R 1210  
 RESULT 8  
 POLG HCVJ8 STANDARD; PRT; 3033 AA.  
 AC P26651;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide (Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP1) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus).  
 -----





DT	01-DEC-1992 (Rel. 24, Created)
DT	01-DEC-1992 (Rel. 24, Last sequence update)
DT	01-FEB-1994 (Rel. 28, Last annotation update)
DS	Structural protein 2 precursor (ORF2).
OC	Hepatitis E virus (strain Burma) (HEV).
OC	Viruses; ssRNA positive-strand viruses, no DNA stage;
OC	Hepatitis E-like viruses.
OX	NCBI_TaxID=31767;
RP	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=92024067; PubMed=1926770;
RA	Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA	Fry K.E., Reyes G.R.;
RT	"Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT	full-length viral genome";
RL	Virology 185:120-131(1991).
CC	-I- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC	THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC	BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; M73218; AAA45736.1; -.
DR	PIR; C40778; VHWMH2.
DR	InterPro; IPR004261; SP2.
DR	InterPro; IPR008975; Viral_cap_coat.
DR	Pfam; PF03014; SP2; 1.
KM	Signal.
FT	SIGNAL.
FT	CHAIN
SO	SEQUENCE 660 AA; 70978 MW; 5832A013CCCA61C CRC64;
	POTENTIAL.
	STRUCTURAL PROTEIN 2.
	Query Match
	Best Local Similarity 19.0%; Score 93; DB 1; Length 660;
	Matches 72; Conservative 47; Mismatches 119; Indels 140; Gaps 16;
QY	12 AAGIIRACMLVRRKAGGHVQMAFMKLAALGTIVYDHLPLQDMAAGLRDLAVNEPV 71
DB	188 ARATIRYRPLVPMNAVGVYASISFMPQTTPPTVS-----DNMSITSDVAILVQPG 239
QY	72 IFSDEMKIITWCAADTACGDIISGLFVRSARGRHILGPAD--NFGQGGRLAPI--TA 128
DB	240 IASLVL-----FSERLHYANQGMRSVETSGVA 267
QY	129 YSQQTRELL-----GCITTSLTG-----146
DB	268 EEEATSGIWLMLCIHGSJLMSYNTFPYCALGLDFALELEFRNLTPGNTTRVSRVSTA 327
QY	147 --RKAKVVEGVCVVSFATQSFIA-----TCVNG-----CMTVEH-----180
DB	328 RHRIRRRADGTAELETTAATREPKDLFTSTNGVGEIGRIALTLEFLAATLLGSLPTEL 387
QY	181 --GAG-----SKTLAGEPKG--PITOMYTNVDODLVGWQAPGASMTBCTGSSDLYLV-- 230
DB	388 ISSAGCOLFYSRPVSNANGSPVTLVLYTSVENA-----QQDGKGIAPHDIDIGESRVVIQDY 443
QY	231 --TRHADVIPIRRRGDSNG--SLSSRPVSYLK-----GSSGGPLLCPSGHANGIF 277
DB	444 DNGHEQCRFPSPAPSPFSLVRANDVLMWLSLTAAYDOSTYSGSTGEPVYV--SDSVTLV 501
QY	278 RAAVCTRGVAKAVDFIFV 295
DB	502 NVATGAQAVARSLDWTKV 519

ID VST2 HEVPA STANDARD; PRT; 660 AA.  
AC P33426;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Structural protein 2 precursor (ORF2).  
OS Hepatitis E virus (strain Pakistan) (HEV).  
OC Viruses; ssRNA positive-strand viruses; no DNA stage;  
OC Hepatitis E-like viruses.  
OC NCBI TaxID=33774;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92115700; Pubmed=1731327;  
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,  
RA Malik I.A., Iqbal M., Purcell R.H.;  
RT "Characterization of a prototype strain of hepatitis E virus."  
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M80581; AAA45727.1; -  
DR InterPro; IPR004261; SP2.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF03014; SP2; 1.  
KW Signal.  
FT SIGNAL 1 22 BY SIMILARITY  
FT CHAIN 23 660 STRUCTURAL PROTEIN 2. CRC64;  
SQ SEQUENCE 660 AA; 70980 MW; 80858C53CFB46FD3 CRC64;  
Query Match 5.9%; Score 93; DB 1; Length 660;  
Best Local Similarity 19.0%; Pred. No. 2.3;  
Matches 72; Conservative 49; Mismatches 117; Indels 140; Gaps 17;  
QY 12 AGLIRACMLVRKAGHYQVAFKALALTYVVDHLTPLODMAHAGRLDAVAVERV 71  
DB 188 ARATIRYRPVLPVNAVGVATISIFWPQTITPTSV-----DNMSITSDVRLVQPG 239  
QY 72 IFSPMEKILTWGMDTAACDIISGLVARSRGREILGPAD--NFEQGMRL----- 123  
DB 240 IASLEVI-----PSRLHTRNQMKRSEVTSQVA 267  
QY 124 -----ADITAYSOQT-RGLIGCI-----ITSLTGRDKNQ----- 151  
DB 268 EEEATSGVLMCIHGSPVNSVTPTYGALGLDPALELEFRMLTGTGNTTRVSRYSSTA 327  
QY 152 -----VEGEVQVVSATQSEFLA-----TCVNGV-----CMTVFP----- 180  
DB 328 RHRIRAGADGTAEITTAATRFMKDLVFTSTNGVGEIGRAIALTLRLADLTGLPTEL 387  
QY 181 --GAG-----SKTLAGEPKG-PITOMYTNVDOLVGMQAPGASMTPTCGSSDLYLV-- 230  
DB 388 ISSAGGLLFYRSPVANSGETVLYTSVENA-----QODGIAIPHDIDGESRVVIOXY 443  
QY 231 -TRHADVIPTARRRDSRG-SLSRPVSYLK-----SSGGPILCPGHAVGIF 277  
DB 444 DNGHEQDRPTPSPAPSRPFSVLRANDVLMSTLAEYDQSTGYSSSTGPPVYV--SDSVTLV 501  
QY 278 RAAVCTRGVAKAVFIPV 295  
DB 502 NVATGAQAVARSLDMTRY 519

ID TRFE HORSE STANDARD; PRT; 706 AA.  
AC P27425;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Seroctransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal  
DE binding globulin).  
GN TF.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OC NCBI TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93277958; Pubmed=8504171;  
RA Carpenter M.A., Broad T.E.;  
RT "The cDNA sequence of horse transferrin."  
RL Biochim. Biophys. Acta 1173:230-232(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Extraembryonic tissue;  
RA McDowell K.J., Adams M.H., Baker C.B.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.  
CC -1- DOMAIN: Composed of two homologous domains.  
CC -1- SIMILARITY: Belongs to the transferrin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M69020; AAA30958.1; -  
DR EMBL; U21127; AAA63684.1; -  
DR PIR; S33761; S33761.  
DR HSP; P02787; IABE.  
DR InterPro; IPR001156; Transferrin.  
DR Pfam; PF00405; transferrin; 2.  
DR PRINTS; PR00422; TRANSFERRIN.  
DR SMART; SM00094; TR\_FER\_2.  
DR PROSITE; PS00205; TRANSFERRIN\_1; 2.  
DR PROSITE; PS00206; TRANSFERRIN\_2; 2.  
DR PROSITE; PS00207; TRANSFERRIN\_3; 2.  
KW Transferrin; Iron transport; Glycoprotein; Metal-binding; Repeat;  
KW Signal.  
FT SIGNAL 1 19 BY SIMILARITY  
FT CHAIN 20 706 SEROTRANSFERRIN.  
FT REPEAT 20 357 1.  
FT REPEAT 358 706 2.  
FT DISULFID 26 64 BY SIMILARITY.  
FT DISULFID 36 55 BY SIMILARITY.  
FT DISULFID 134 215 BY SIMILARITY.  
FT DISULFID 174 190 BY SIMILARITY.  
FT DISULFID 177 198 BY SIMILARITY.  
FT DISULFID 187 200 BY SIMILARITY.  
FT DISULFID 248 262 BY SIMILARITY.  
FT DISULFID 360 623 BY SIMILARITY.  
FT DISULFID 366 398 BY SIMILARITY.  
FT DISULFID 376 389 BY SIMILARITY.  
FT DISULFID 423 701 BY SIMILARITY.  
FT DISULFID 441 664 BY SIMILARITY.  
FT DISULFID 474 550 BY SIMILARITY.

RESULT 12  
TRFE\_HORSE

```

FT DISULFID 498 692 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.
FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (BY SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409DBA CRC64;

```

```

Query Match 5.8%; Score 92.5; DB 1; Length 706;
Best Local Similarity 21.5%; Pred. No. 2.8;
Matches 64; Conservative 43; Mismatches 109; Indels 81; Gaps 17;

```

```

QY 33 MAEKLALNGTYV---YDHLTPLOQVNAHGLRLAVAVEPVSDEVKIITMGA-----85
DB 321 LGFLRIPADPTWLYGLEYVT-----AIRNLREDIPBPKD-ECKKVMKALGH 371
QY 86 DTAACGD-IISGLPVASRRGR-----EILGPDNFEQGQWRL-----LAPITAY 129
DB 372 EKYKCEWVSNGGNIACESAQSTEDCIARKVGEADAMSLDGFYIAGKGLVPLAE 431
QY 130 SQQTRGLGCIITSLTRDKNQVGEVQVSTATQSLATCVAGVCTVTHGASKTLAG 189
DB 432 NYEIRSSGACVDPEESYH-----AAVAVKSSDDPLT-----W-----NSLKG 470
QY 190 PKSPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLYLVTHHADVIPRRRGDSRSL 249
DB 471 KK-----SCHTGVDR-TAGMNI PMGL-----LYSEIHGCEPDKFRREGCAPGR 513
QY 250 LSRPRVSYLKSSGGR-LIC-PSGHA-----VGIIFRAVCTRGVAKAVDFIPVESWE 299
DB 514 RNSTLCLNLGASGPGRECEPNNHERRYGYTGAFLVLEKGDVA---FVKQATVE 566

```

```

RESULT 13
VST2_HEVME STANDARD; PRT; 659 AA.

```

```

AC 003500;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor.
OS Hepatitis E virus (strain Mexico) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=31768;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=3079857; Pubmed=1448913;
RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,
RA Bradley D.W., Tam A.W., Reyes G.R.;
RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
RT E virus (HEV)."
RL Virology 191:550-558(1992).

```

```

CC -I- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74506; AAA45732.1; -
DR PIR; B44212; B44212.
DR InterPro; IPR004261; SP2.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF03014; SP2; 1.
KM Signal.
FT CHAIN 1 22 BY SIMILARITY.
FT SIGNAL 23 659 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 659 AA; 70640 MW; CFF5B75BFD8BE2C CRC64;

```

```

Query Match 5.6%; Score 89.5; DB 1; Length 659;
Best Local Similarity 17.8%; Pred. No. 4.7;
Matches 67; Conservative 50; Mismatches 121; Indels 139; Gaps 14;

```

```

QY 12 AGLIRACLVYKAKAGHYVQKAEKALALTYVDHLTPLOQVNAHGLRLAVAVEPV 71
DB 188 AARTIRYRELVNNAVGAISISFWPQTTPTSV-----DNMSITSDVRLVDPG 239
QY 72 IFSDMEVKIITWGAADTAACGDIISGLPVASRRGRILLGPAD--NFGQGWRL-----123
DB 240 ISELVI-----APTANSQ-----QTRGLGCIITSLTGR-----147
QY 124 -----APTANSQ-----QTRGLGCIITSLTGR-----147
DB 268 EEEATSGVLMLCINGSPVNSYNTPTGTALGLDPALIEFANLITQVNTVRSYSSTA 327
QY 148 --DKQVGEVQVNSTATQSFLLA-----TCVGV-----CWTVHGASKTL 187
DB 328 RNSARGADSTALTLTATATRFKDLHFTGLNGVGVGRIALLTLNADTLIGLPTLEI 387
QY 188 AGPKG-----PTQMTNVDDLVGMQAPPGARSMTPTCGSSDLYLV-----230
DB 388 SSAGCQLFYSRPVVSANGEPYKLYTSVENA-----QDQKGAIPHIDLDGSRVVIQDYD 443
QY 231 TTHADVIPRRRGDSRG-SLSRPVSYLK-----GSSGGLICSGHAYGIFR 278
DB 444 NQHEODRFPSPAPRPSPVLANVDVLSLTAAYDOSTYGSSTGSPYIT--SDSVTLVN 501
QY 279 AAVCTRGVAKAVDFIPV 295
DB 502 VATGAQAVARSLDMKV 518

```

```

RESULT 14
POLG_TBENV STANDARD; PRT; 3414 AA.
AC P14336; O88493;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE (NS5)].
OS Tick-borne encephalitis virus (Western subtype) (TBEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Flavivirus.
OX NCBI_TaxID=11088;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS.

```



```

FT TURN 414 416
FT STRAND 418 425
FT TURN 433 434
FT TURN 438 439
FT STRAND 441 446
FT TURN 447 448
FT STRAND 449 449
FT STRAND 451 455
FT HELIX 457 459
FT STRAND 460 467
FT HELIX 468 470
FT TURN 475 476
FT STRAND 477 481
FT TURN 484 485
FT TURN 487 488
FT STRAND 492 496
FT HELIX 497 501
FT TURN 502 502
FT STRAND 507 508
FT TURN 510 511
FT STRAND 516 516
FT TURN 517 517
FT HELIX 518 521
FT STRAND 522 524
FT STRAND 528 528
FT TURN 529 530
FT STRAND 531 531
FT STRAND 534 536
FT HELIX 541 547
FT TURN 548 550
FT STRAND 553 557
FT TURN 558 559
FT STRAND 560 562
FT STRAND 567 573
FT TURN 575 576
FT STRAND 577 577
FT TURN 582 583
FT STRAND 586 586

```

Query Match 5.5%; Score 88; DB 1; Length 3414;

Best Local Similarity 23.0%; Pred. No. 47;

Matches 59; Conservative 28; Mismatches 86; Indels 84; Gaps 12;

```

QY 49 HLTPLD-----WALAGRLDAVAVPVIFFSMERKITTGADTACGDIISGLPVSA 101
DB 1442 HLTLEKEBERMAFWL1AGL-----AASAIHWSGILGWGLWLTETLRSSRRSDLVFSG 1496
QY 102 RRGREIILGPADNFGGQWRLLAPITAYSCQTRGLGCIITSLTRDKNQVEGEVQVST 161
DB 1497 QGGERDRPREVDGV-YRIFSGLFMGQ-----NQVG-----VGY 1532
QY 162 ATGCFATCNGVGVTFPHGAG--SKTLAGEKCPITQMTNVNDQIV-----GW 208
DB 1533 GSKGVLTHT-----MMHYTRGALSIDAVAGP-----YADVAEDVVCYGAMSLSEKN 1581
QY 209 QA-----PPGARSMTPCTCGSSDLYLVTRHADVIPIVRRGRDSRGLSLSPRVSYLKG 260
DB 1582 KGEIVQVHAFFPG-RAHEVHOCGCELLIDT-----GRKIGALPIDLVKG 1625
QY 261 SSGPPLCPSGHVAIVG 277
DB 1626 TSGSPILNAQGVVGLY 1642

```

RESULT 15  
 POLG\_TBEVS STANDARD; PRT; 3412 AA.  
 AC P07720; P07721; Q88475; Q88476; Q88477; Q88478; Q88479; Q88877;  
 AC Q88878; Q88879;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Genome polypeptide (Contains: Capsid protein C (Core protein); Matrix

DE protein (Envelope protein M); Major envelope protein E; Nonstructural  
 DE proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase  
 DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)  
 DE (NS5).  
 OS Tick-borne encephalitis virus (strain Sofjin) (TBEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_TaxId=11087;  
 RN [1]  
 RP MEDLINE=90101381; PubMed=2136778;  
 RA Pletnev A.G., Yamshchikov V.F., Blinov V.M.;  
 RT "Nucleotide sequence of the genome and complete amino acid sequence  
 of the polyprotein of tick-borne encephalitis virus.";  
 RL Virology 174:250-263(1990).  
 RN [2]  
 RP SEQUENCE OF 1-1190 FROM N.A.  
 RX MEDLINE=88319988; PubMed=2970626;  
 RA Yamshchikov V.F., Pletnev A.G.;  
 RT "Nucleotide sequence of the genome region encoding the structural  
 proteins and the NS1 protein of the tick borne encephalitis virus.";  
 RL Nucleic Acids Res. 16:7750-7750(1988).  
 RN [3]  
 RP SEQUENCE OF 1-683 AND 777-1002 FROM N.A.  
 RX MEDLINE=86220766; PubMed=3709796;  
 RA Pletnev A.G., Yamshchikov V.F., Blinov V.M.;  
 RT "Tick-borne encephalitis virus genome. The nucleotide sequence coding  
 for virion structural proteins.";  
 RL FEBS Lett. 200:317-321(1986).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 hydrophobic, suggesting a possible membrane-related function.  
 CC NS3 and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 precursor polyprotein, commonly with Asp or Glu in the P6  
 position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 {RNA} (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and RNA.  
 CC -1- MISCELLANEOUS: The nonstructural protein NS1 presents two  
 alternative cleavage sites for its C-terminus, which may define a  
 soluble or a membrane-bound form of NS1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X07755; CAA30581.1; -  
 DR EMBL; X03870; CAA27500.1; -  
 DR EMBL; X03870; CAA27501.1; ALT SEQ.  
 DR EMBL; X03870; CAA27502.1; ALT SEQ.  
 DR EMBL; X03870; CAA27503.1; ALT SEQ.  
 DR EMBL; X03871; CAA27505.1; -  
 DR MEROPS; S07.006; -  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR000336; Flavi\_glycoprote.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR011157; Flavi\_NS1.  
 DR InterPro; IPR000752; Flavi\_NS2A.  
 DR InterPro; IPR000487; Flavi\_NS2B.  
 DR InterPro; IPR000404; Flavi\_NS4A.  
 DR InterPro; IPR001528; Flavi\_NS4B.  
 DR InterPro; IPR002028; Flavi\_NS5.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR01650; Helicase\_C.

DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR001850; peptidase S7.  
 DR InterPro: IPR007095; RNA\_POL\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_POL\_PSVIT.  
 DR InterPro: IPR002877; Rnuv\_FtsJ.  
 DR Pfam: PF01003; Flavi\_capsid.1.  
 DR Pfam: PF02832; Flavi\_glycop\_C.1.  
 DR Pfam: PF00869; Flavi\_glycoprot.1.  
 DR Pfam: PF00949; Flavi\_helicase.1.  
 DR Pfam: PF01004; Flavi\_M.1.  
 DR Pfam: PF00948; Flavi\_NS1.1.  
 DR Pfam: PF01005; Flavi\_NS2A.1.  
 DR Pfam: PF01002; Flavi\_NS2B.1.  
 DR Pfam: PF01350; Flavi\_NS4A.1.  
 DR Pfam: PF01349; Flavi\_NS4B.1.  
 DR Pfam: PF00972; Flavi\_NS5.1.  
 DR Pfam: PF01570; Flavi\_propep.1.  
 DR Pfam: PF01728; FtsJ.1.  
 DR Pfam: PF00271; helicase\_C.1.  
 DR ProDom: PD001556; Flavi\_glycoprote.1.  
 DR ProDom: PD001496; Flavi\_NS1.1.  
 DR PROSITE: PS00690; DEAH ATP HELICASE; FALSE NEG.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Hydrolyase; Helicase;  
 KW ATP-binding; Transmembrane; Nonstructural protein.  
 FT INIT\_MET 1  
 FT CHAIN 1 112  
 FT PROPEP 113 205  
 FT CHAIN 206 280  
 FT CHAIN 281 776  
 FT CHAIN 777 71128  
 FT CHAIN 71129 1358  
 FT CHAIN 1359 1489  
 FT CHAIN 1490 2110  
 FT CHAIN 2111 2259  
 FT CHAIN 2260 2510  
 FT CHAIN 2511 3412  
 FT NP\_BIND 1688 1695  
 FT SITE 1779 1782  
 FT TRANSMEM 101 112  
 FT TRANSMEM 247 259  
 FT TRANSMEM 266 280  
 FT TRANSMEM 738 751  
 FT DISULFID 283 310  
 FT DISULFID 340 396  
 FT DISULFID 354 385  
 FT DISULFID 372 401  
 FT DISULFID 466 570  
 FT DISULFID 587 618  
 FT CARBOHYD 144 144  
 FT CARBOHYD 434 434  
 FT CARBOHYD 861 861  
 FT CARBOHYD 983 983  
 FT CARBOHYD 999 999  
 FT CARBOHYD 1228 1228  
 FT CARBOHYD 2447 2447  
 FT CARBOHYD 2466 2466  
 FT CONFLICT 381 381  
 FT CONFLICT 850 850  
 SQ SEQUENCE 3412 AA; 37976 MW; 0F61CE6DCDC5965 CRC64;

Query Match 5.5%; Score 87; DB 1; Length 3412;

Best Local Similarity 24.4%; Pred. No. 58; Mismatches 58; Indels 74; Gaps 11;

Db 50; Conservative 23; Matches 58; Indels 74; Gaps 11;  
 QY 95 SGLPVSARRGRREILLGPADNFEQGGWELLAPITAYSQOTRGLGCIITSLTGRDKNQVHG 154  
 1490 SGLVFSGGGRERDRPFVYDGV-YRIFSP-----GLLM-----G 1524  
 QY 155 EVQV-VSTAQSPLATCVNGVCVTFVHGAG---SKTLAGEKXPITOMYTNVDQDIY----- 206  
 1525 QRQVGVGYGSKGVLT-----MMHYTFGALSIDDAVAGP-----YMAADVKEVVCYGG 1573

QY 207 -----GWA-----PGARSMPTCGSDLYVTRHADVIYVRRRGDSRGLISP 252  
 Db 1574 AMSLEKRRKGETVQVHAFFPG-RHBEVHQCPGSLLDLT-----GRRIGA 1617  
 QY 253 RPSVYLKSSGGPILCPB8GAVGIF 277  
 Db 1618 VPIDAKTSGSPILNSQGVVGLY 1642

Search completed: May 6, 2004, 09:31:49  
 Job time : 7.32568 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 28.339 Seconds

(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-10  
Perfect score: 1589  
Sequence: 1 AGITKVPYFRAQGLRACM.....RGVAKAVDFIPVSMETMR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriopl:\*
- 17: sp\_archaeopl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1550	97.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
2	1545	97.2	3010	12 Q68826	Q68826 hepatitis c
3	1545	97.2	3010	12 P90191	P90191 hepatitis c
4	1544	97.2	3010	12 Q9DRE6	Q9DRE6 hepatitis c
5	1542	97.0	3010	12 Q9DRE4	Q9DRE4 hepatitis c
6	1541	97.0	3010	12 Q9DRE6	Q9DRE6 hepatitis c
7	1540	96.9	3010	12 P88603	P88603 hepatitis c
8	1539	96.9	3010	12 Q9J3H5	Q9J3H5 hepatitis c
9	1539	96.9	3010	12 Q807P3	Q807P3 hepatitis c
10	1537	96.7	3010	12 Q9J3F9	Q9J3F9 hepatitis c
11	1534	96.5	361	12 Q70815	Q70815 hepatitis c
12	1534	96.5	3008	12 Q9J3P4	Q9J3P4 hepatitis c
13	1534	96.5	3010	12 Q9J3H3	Q9J3H3 hepatitis c
14	1533	96.5	3010	12 Q9J3H2	Q9J3H2 hepatitis c
15	1533	96.5	3010	12 Q9J3I0	Q9J3I0 hepatitis c
16	1532	96.4	3010	12 Q9J3Y3	Q9J3Y3 hepatitis c

17	1532	96.4	3010	12 Q9J3H6	Q9J3H6 hepatitis c
18	1531	96.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
19	1531	96.3	3010	12 Q9J3H8	Q9J3H8 hepatitis c
20	1531	96.3	3010	12 Q9J3H9	Q9J3H9 hepatitis c
21	1531	96.3	3010	12 Q9J3H5	Q9J3H5 hepatitis c
22	1531	96.3	3010	12 Q9J3H6	Q9J3H6 hepatitis c
23	1530	96.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
24	1530	96.3	3010	12 Q9J3H8	Q9J3H8 hepatitis c
25	1530	96.3	3010	12 Q9J3H9	Q9J3H9 hepatitis c
26	1529	96.2	3013	12 Q9J3H4	Q9J3H4 hepatitis c
27	1528	96.2	1186	12 Q81755	Q81755 hepatitis c
28	1528	96.2	2284	12 Q81817	Q81817 hepatitis c
29	1528	96.2	3010	12 Q68788	Q68788 hepatitis c
30	1528	96.2	3010	12 P89965	P89965 hepatitis c
31	1528	96.2	3010	12 Q9DRE7	Q9DRE7 hepatitis c
32	1528	96.2	3010	12 Q9DRE8	Q9DRE8 hepatitis c
33	1528	96.2	3014	12 Q9DRE9	Q9DRE9 hepatitis c
34	1527	96.1	3010	12 Q9DRE2	Q9DRE2 hepatitis c
35	1527	96.1	3011	12 Q9DRE3	Q9DRE3 hepatitis c
36	1526	96.0	361	12 Q70818	Q70818 hepatitis c
37	1526	96.0	3010	12 Q9J3H6	Q9J3H6 hepatitis c
38	1526	96.0	3010	12 Q9J3H7	Q9J3H7 hepatitis c
39	1525	96.0	3010	12 Q9J3H8	Q9J3H8 hepatitis c
40	1524	95.9	3010	12 Q9J3H9	Q9J3H9 hepatitis c
41	1523	95.8	3010	12 Q9J3H5	Q9J3H5 hepatitis c
42	1523	95.8	3010	12 Q9J3H6	Q9J3H6 hepatitis c
43	1523	95.8	3010	12 Q9J3H7	Q9J3H7 hepatitis c
44	1523	95.8	3010	12 Q9J3H8	Q9J3H8 hepatitis c
45	1522	95.8	3010	12 Q9J3H9	Q9J3H9 hepatitis c

#### ALIGNMENTS

RESULT 1

Q9J3H7 PRELIMINARY, PRT, 3010 AA.

AC Q9J3H7;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_Taxid=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MD15;  
RA Nagayama K., Kurotsaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease progression."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBMITTER: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL; AF207756; AAF65946.1; -  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.  
DR PIR; P00804; P00804.  
DR PIR; P80329; P80329.  
DR HSSP; P26663; IJXP.  
DR GO; GO:016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F:electon transporter activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.



Matches 292; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITVPYFVPAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 904 AGITAVPFVPAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 QY 61 LRDIAVAEVPYFSPMEVKITTWGADTAACDIIISGLPVSRARRREIILGPDNFEQGW 120  
 Db 964 LRDIAVAEVPYFSPMEVKITTWGADTAACDIIISGLPVSRARRREIILGPDNFEQGW 1023  
 QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMWTFH 180  
 Db 1024 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMWTFH 1083  
 QY 181 GAGSKTLAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLVYTRHADVIPVR 240  
 Db 1084 GAGSKTLAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLVYTRHADVIPVR 1143  
 QY 241 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 300  
 Db 1144 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
 QY 301 TMR 303  
 Db 1204 TMR 1206

## RESULT 3

P90191 PRELIMINARY; PRT; 3010 AA.

AC P90191; PRT; 3010 AA.  
 DT 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RA Yamanoto C., Izumi N., Marumo F., Seto C.,  
 RA "Comparison of full-length sequences of interferon-sensitive and  
 RT resistant hepatitis C virus 1b.";  
 RL J. Clin. Invest. 96:224-230(1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: D50482; ABA05073.1; -;  
 DR PIR: A61196; A61196.  
 DR PIR: P00254; P00254.  
 DR PIR: P00804; P00804.  
 DR PIR: P80329; P80329.  
 DR PIR: 1DXY: 12-JAN-01.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:000524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003668; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.

DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser\_tyrpsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4.  
 DR InterPro: IPR001480; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5A.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR004109; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5A; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC1; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; transferase; transmembrane.  
 FT CHAIN 1 191  
 FT CHAIN 192 383  
 FT CHAIN 384 809  
 FT CHAIN 810 1026  
 FT CHAIN 1027 1657  
 FT CHAIN 1658 1771  
 FT CHAIN 1772 1972  
 FT CHAIN 1973 2419  
 FT CHAIN 2420 3010  
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match 97.2%; Score 1545; DB 12; Length 3010;

Best Local Similarity 96.0%; Pred. No. 6,1e-127;

Matches 291; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGITVPYFVPAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 904 AGITAVPFVPAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 QY 61 LRDIAVAEVPYFSPMEVKITTWGADTAACDIIISGLPVSRARRREIILGPDNFEQGW 120  
 Db 964 LRDIAVAEVPYFSPMEVKITTWGADTAACDIIISGLPVSRARRREIILGPDNFEQGW 1023  
 QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMWTFH 180  
 Db 1024 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMWTFH 1083  
 QY 181 GAGSKTLAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLVYTRHADVIPVR 240  
 Db 1084 GAGSKTLAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLVYTRHADVIPVR 1143  
 QY 241 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 300  
 Db 1144 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
 QY 301 TMR 303  
 Db 1204 TMR 1206

RESULT 4  
ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
AC Q9DTE6;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HCV142;  
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
RA Mishiro S.;  
RT Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
RT with hepatocellular carcinoma: the 'progression score' revisited.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA (BY SIMILARITY).  
DR EMBL; AB049091; BAB18804.1; -.  
DR PIR; A61196; A61196.  
DR PIR; PS0329; PS0329.  
DR HSBP; P26663; IUXP.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0006118; F:electron transport; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR009003; Cys\_Ser\_typsin.  
DR InterPro; IPR00345; CysC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002511; HCV\_NSI.  
DR InterPro; IPR002518; HCV\_NSI.  
DR InterPro; IPR00745; HCV\_NS2.  
DR InterPro; IPR001490; HCV\_NS4a.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_NS5a.  
DR InterPro; IPR001650; HCV\_NS5a.  
DR InterPro; IPR004109; peptidase\_C29.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01543; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NSI; 1.  
DR Pfam; PF01538; HCV\_NS1; 1.  
DR Pfam; PF02907; HCV\_NS2; 1.  
DR Pfam; PF01506; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS5a; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_Rdrp; 1.  
DR ProDom; PD186062; HCV\_NSI; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; 1.  
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KW Hydroxylase; Nonstructural protein; Polypeptide;  
KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;  
Query Match 97.2%; Score 1544; DB 12; Length 3010;  
Best Local Similarity 95.7%; Pred. No. 7,4e-127;  
Matches 290; Conservative 8; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AGITKPYVFAAGLILACMLVRAAGHYVQAEMKALATGYVYDHLTPIDQVMAHAG 60  
DB 904 AGITRVFPVFAAGLILACMLVRAAGHYVQAEMKALATGYVYDHLTPIDQVMAHAG 963  
QY 61 LRDIAVAVEPVFSDMEVKIITWGDADTAACGDIISGLPVASRRGREIILGADNFEQGV 120  
DB 964 LRDIAVAVEPVFSDMETKIITWGDADTAACGDIISGLPVASRRGREIILGADNFEQGV 1023  
QY 121 RLAPITAYSOQTGILGCIITSLTGSDKNQVGEVGVVSTATQSFATCNGVCWTFH 180  
DB 1024 RLAPITAYSOQTGILGCIITSLTGSDKNQVGEVGVVSTATQSFATCNGVCWTFH 1083  
QY 181 GAGSKTLAAGKPIITOMTWNVDOLVGMQAPPGARSMTPTCCSSDLYLTRADVI 240  
DB 1084 GAGSKTLAAGKPIITOMTWNVDOLVGMQAPPGARSMTPTCCSSDLYLTRADVI 1143  
QY 241 RRGDSRGLSPRVSYLKGSSGSPLLCPGSHAVGIFRAAVCTRGKAVDFIPVSEMET 300  
DB 1144 RRGDSRGLSPRVSYLKGSSGSPLLCPGSHAVGIFRAAVCTRGKAVDFIPVSEMET 1203  
QY 301 TWR 303  
DB 1204 TWR 1206  
RESULT 5  
ID Q9DTE4 PRELIMINARY; PRT; 3010 AA.  
AC Q9DTE4;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HCV150;  
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
RA Mishiro S.;  
RT Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
RT with hepatocellular carcinoma: the 'progression score' revisited.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA (BY SIMILARITY).  
DR EMBL; AB049093; BAB18806.1; -.  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.  
DR PIR; P00804; P00804.  
DR PIR; PS0329; PS0329.  
DR HSBP; P26663; IUXP.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.

DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0006740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; Cys\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR004109; Peptidase C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF01506; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase C; 1.  
 DR Pfam: PF00998; viral RdRp; 1.  
 DR Pfam: PF01806; HCV\_NS1; 1.  
 DR SMART: SM00467; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 KM Coat protein; Envelope protein; Nonstructural protein;  
 Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327324 MW; 3DE6CF249BD151C CRC64;

Query Match 97.0%; Score 1542; DB 12; Length 3010;  
 Best Local Similarity 95.7%; Pred. No. 1,1e-126;  
 Matches 290; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVFYFRAOGLIRACMLVTRKAGGHVYVMAFMKLALITGVYVDHLTPLODMAHAG 60  
 |||||  
 DB 904 AGITKVFYFRAOGLIRACMLVTRKAGGHVYVMAFMKLALITGVYVDHLTPLODMAHAGS 963  
 |||||  
 QY 61 LRDLAAVEPVITSDMEVKITTWGADTAACDIIISGLPVASARRGELLIGPADNFGQGN 120  
 |||||  
 DB 964 LRDLAAVEPVITSDMEVKITTWGADTAACDIIISGLPVASARRGELLIGPADNFGQGN 1023  
 |||||  
 QY 121 RLAPITAYSGOQRGLIGCIITSLTGRDNQVEGVEVYVSTAROSPLATGVNVCMTVTH 180  
 |||||  
 DB 1024 RLAPITAYSGOQRGLIGCIITSLTGRDNQVEGVEVYVSTAROSPLATGVNVCMTVTH 1083  
 |||||  
 QY 181 GAGSKTLAGEKGPITQYTVTVDDLVGMQAPPGARSMTPTCTGSSDLYLTVTRADVI 240  
 |||||  
 DB 1084 GAGSKTLAGEKGPITQYTVTVDDLVGMQAPPGARSMTPTCTGSSDLYLTVTRADVI 1143  
 |||||  
 QY 241 RRGDSRGSLSPPRVSTLKSSSGGFLICSGHAGVIFRAVCTRGVAKAVDFPVESMET 300  
 |||||  
 DB 1144 RRGDSRGSLSPPRVSTLKSSSGGFLICSGHAGVIFRAVCTRGVAKAVDFPVESMET 1203  
 |||||  
 QY 301 TMR 303  
 |||||

DB 1204 TMR 1206  
 RESULT 6  
 ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTE6;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxId=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV221;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 Hatanata T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 with hepatocellular carcinoma: the 'progression score' revisited";  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC CC EMBL; AB049101; BAB1814.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00329; P00329.  
 DR HSSP; P26663; LUXP.  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:000489; F:electron transporter activity; IEA.  
 DR GO:0016787; F:hydrolase activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; Cys\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR004109; Peptidase C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.

DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; helicase C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR Pfam: PF0186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC 1.  
 DR PROSITE; PS00180; CYTOCHROME C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM ATP-binding; Nonstructural protein; Polyprotein;  
 KM Hydroxylase; Nonstructural protein; Polyprotein;  
 KM RNA-directed RNA polymerase; Transferrase;  
 SQ SEQUENCE 3010 AA; 327108 MW; DE182D810EF78EE4 CRC64;

Query Match 97.0%; Score 1541; DB 12; Length 3010;  
 Best Local Similarity 96.0%; Pred. No. 1.4e-126;  
 Matches 291; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKYVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMWANG 60  
 DB 904 AVILKPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMWANG 963

QY 61 LRLDAVAVPEVIFSMEVKIITWGADTAACGDIISGLPVSARRGRILLGPADSEEGGW 120  
 DB 964 LRLDAVAVPEVIFSMEVKIITWGADTAACGDIISGLPVSARRGRILLGPADSEEGGW 1023

QY 121 RLAPITAYVSOOTRGLGCIITSLTGRDKVGEVGVVSTATQSEFLATCVNGVCWTVF 180  
 DB 1024 RLAPITAYVSOOTRGLGCIITSLTGRDKVGEVGVVSTATQSEFLATCVNGVCWTVF 1083

QY 181 GAGSKTLGPKGPITQMTNVDOLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPIR 240  
 DB 1084 GAGSKTLGPKGPITQMTNVDOLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPIR 1143

QY 241 RRGDSRSILSPRVSYLKSSGGPILCPGSHVGFRAVCTRGVAKAVDFIPVSMET 300  
 DB 1144 RRGDSRSILSPRVSYLKSSGGPILCPGSHVGFRAVCTRGVAKAVDFIPVSMET 1203

QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 7  
 P88803 PRELIMINARY; PRT; 3010 AA.  
 ID P88803;  
 AC P88803;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome Polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RA Enomoto N.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RA MEDLINE=75340824; PubMed=7542279;  
 RX Enomoto N., Sakuma T., Asahina Y., Kurosaki M., Murakami T.,  
 RA Yamamoto C., Izumi N., Martomo F., Sato C.;  
 RT "Comparison of full-length sequences of interferon-sensitive and  
 RT resistant hepatitis C virus 1b."  
 RU J. Clin. Invest. 96:224-230 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MNNA (BY SIMILARITY).  
 DR EMBL; D50484; BAA09075.1; -  
 DR PIR; A61196; A61196.

DR HSPB; P26663; INS3.  
 DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F: structural molecule activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: translation activity; IEA.  
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_tyrpsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002519; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR Pfam; PF0186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 FT CHAIN 1  
 FT CHAIN 191  
 FT CHAIN 192 383  
 FT CHAIN 384 809  
 FT CHAIN 810 1026  
 FT CHAIN 1027 1557  
 FT CHAIN 1658 1711  
 FT CHAIN 1712 1972  
 FT CHAIN 1973 2419  
 FT CHAIN 2420 3010  
 SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783FEF88 CRC64;

Query Match 96.9%; Score 1540; DB 12; Length 3010;  
 Best Local Similarity 95.4%; Pred. No. 1.7e-126;  
 Matches 289; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGITKYVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMWANG 60  
 DB 904 AVILKPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMWANG 963

QY 61 LRLDAVAVPEVIFSMEVKIITWGADTAACGDIISGLPVSARRGRILLGPADSEEGGW 120  
 DB 964 LRLDAVAVPEVIFSMEVKIITWGADTAACGDIISGLPVSARRGRILLGPADSEEGGW 1023

QY 121 RLAPITAYVSOOTRGLGCIITSLTGRDKVGEVGVVSTATQSEFLATCVNGVCWTVF 180  
 DB 1024 RLAPITAYVSOOTRGLGCIITSLTGRDKVGEVGVVSTATQSEFLATCVNGVCWTVF 1083

QY 181 GAGSKTLGPKGPITQMTNVDOLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPIR 240

```

Db      1084 GAGSKTLAAGPKPIQTOMTNNVDOLVGMQAPPGARSLTPTCGSSDLVYTRHADVTPVR 1143
QY      241 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 300
Db      1144 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 1203
QY      301 TMR 303
Db      1204 TMR 1206

RESULT 8
ID      09J3H5      PRELIMINARY;      PRT; 3010 AA.
AC      09J3H5
DT      01-OCT-2000 (TReMBLrel. 15, Created)
DT      01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MD17;
RA      Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT      "Characteristics of hepatitis C viral genome associated with disease
RT      progression.";
RL      Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND RNA (BY SIMILARITY).
EMBL    AF207758; AAF65948.1; -.
DR      PIR; A61196; A61196.
DR      PIR; P00246; P00246.
DR      PIR; P00254; P00254.
DR      PIR; P50329; P50329.
DR      HSSP; P27958; 1HE1.
DR      GO; GO:0016021; C:integral to membrane; IEA.
DR      GO; GO:0019028; C:viral capsid; IEA.
DR      GO; GO:0019031; C:viral envelope; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR      GO; GO:0005489; F:electron transporter activity; IEA.
DR      GO; GO:0016787; F:hydrolase activity; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0003688; F:RNA-directed RNA polymerase activity; IEA.
DR      GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      GO; GO:0006350; P:transcription; IEA.
DR      GO; GO:0019079; P:viral genome replication; IEA.
DR      GO; GO:0019087; P:viral transformation; IEA.
DR      InterPro; IPR009003; Cys_Ser_cypsin.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR002523; HCV_capsid.
DR      InterPro; IPR002521; HCV_core.
DR      InterPro; IPR002519; HCV_env.
DR      InterPro; IPR002531; HCV_env.
DR      InterPro; IPR002518; HCV_NS1.
DR      InterPro; IPR000745; HCV_NS2.
DR      InterPro; IPR001490; HCV_NS4a.
DR      InterPro; IPR002868; HCV_NS5a.
DR      InterPro; IPR002166; HCV_RdRp.
DR      InterPro; IPR001650; Helicase_C.
DR      InterPro; IPR004109; Peptidase_C29.
DR      InterPro; IPR007095; RNA_pol_DS_PS.

```

```

DR      InterPro; IPR007094; RNA_pol_Psivir.
DR      Pfam; PF01543; HCV_capsid; 1.
DR      Pfam; PF01542; HCV_core; 1.
DR      Pfam; PF01539; HCV_env; 1.
DR      Pfam; PF01560; HCV_NS1; 1.
DR      Pfam; PF01538; HCV_NS2; 1.
DR      Pfam; PF02307; HCV_NS3; 1.
DR      Pfam; PF01006; HCV_NS4a; 1.
DR      Pfam; PF01001; HCV_NS4b; 1.
DR      Pfam; PF01506; HCV_NS5a; 1.
DR      Pfam; PF00271; helicase_C; 1.
DR      Pfam; PF00998; Viral_RdRp; 1.
DR      ProDom; PD186062; HCV_NS1; 1.
DR      SMART; SM00487; DEXDC_1.
DR      PROSITE; PS00130; CYTOCHROME_C_1.
KW      ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW      Hydrolase; Nonstructural protein; Polyprotein;
KW      RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ      SEQUENCE 3010 AA; 326801 MW; 9FEE3D1B9387AA4B CRC64;

Query Match      96.9%; Score 1539; DB 12; Length 3010;
Best Local Similarity 95.7%; Pred. No. 2.1e-126;
Matches 290; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY      1 AGITKVPYFVRAQGHIRACMLVRKAAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAG 60
Db      904 AGITVPIYVRAQGHIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAG 963
QY      61 LRDLAVAEVPVFSMEVKIITWGDYTAACGDIISGLPVASARSGEILIGPADNEGGQW 120
Db      964 LRDLAVAEVPVFSMETKIIITWGDYTAACGDIISGLPVASARSGEILIGPADSDEGGW 1023
QY      121 RLAPITVASQOTRGLGIIITSLGRPNQVEGVQVSTATQSLATCVNVCMTVPH 180
Db      1024 RLAPITVASQOTRGLGIIITSLGRDNQVEGVQVSTATQSLATCINVCMTVPH 1083
QY      181 GAGSKTLAAGPKPIQTOMTNNVDOLVGMQAPPGARSLTPTCGSSDLVYTRHADVTPVR 240
Db      1084 GAGSKTLAAGPKPIQTOMTNNVDOLVGMQAPPGARSLTPTCGSSDLVYTRHADVTPVR 1143
QY      241 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 300
Db      1144 RRGDSRGSLSLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMET 1203
QY      301 TMR 303
Db      1204 TMR 1206

RESULT 9
ID      0807P3      PRELIMINARY;      PRT; 3010 AA.
AC      0807P3;
DT      01-JUN-2003 (TReMBLrel. 24, Created)
DT      01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=WILE;
RA      MEDLINE=22047193; PubMed=12051758;
RA      Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
RA      Nio Y., Hosaka M., Miyazaki Y., Shimotohno K.;
RT      "Subgenomic replicon derived from a cell line infected with the
RT      hepatitis C virus.";
RL      Biochem. Biophys. Res. Commun. 293:993-999(2002).
DR      EMBL; AB080299; BACS4896.1; -.
DR      GO; GO:0019028; C:viral capsid; IEA.
DR      GO; GO:0019031; C:viral envelope; IEA.

```



```

DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008926; F:ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0006118; F:electron transport; IEA.
DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; F:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_Ser_tryptsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR001490; HCV_NS4a.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; Peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_Ps.
DR InterPro: IPR007094; RNA_pol_PsVlr.
DR Pfam: PF01543; HCV_capsid.1.
DR Pfam: PF01542; HCV_core.1.
DR Pfam: PF01539; HCV_env.1.
DR Pfam: PF01560; HCV_NS1.1.
DR Pfam: PF01538; HCV_NS2.1.
DR Pfam: PF02907; HCV_NS3.1.
DR Pfam: PF01006; HCV_NS4a.1.
DR Pfam: PF01001; HCV_NS5a.1.
DR Pfam: PF00271; Helicase_C.1.
DR Pfam: PF00998; Viral_RDRP.1.
DR Pfam: PF0186062; HCV_NS1.1.
DR SMART: SM00487; DEXDC.1.
DR SMART: SM00490; HELICC.1.
DR PROSITE: PS00190; CYTOCHROME_C.1.
DR PolyProtein.
SQ SEQUENCE 3010 AA; 327097 MW; E86418C7A723E686 CRC64;

Query Match 96.3%; Score 1539; DB 12; Length 3010;
Best Local Similarity 96.0%; Pred. No. 2,1e-126;
Matches 291; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

```

```

RESULT 10
ID 09339 PRELIMINARY; PRT; 3010 AA.
AC 09339
DT 01-OCT-2000 (TrEMBLrel.15, Created)
DT 01-OCT-2000 (TrEMBLrel.15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel.25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_taxid=11103;
OX [1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MD3;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
proteins."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND NS5A (BY SIMILARITY).
DR EMBL: AF207774; AAF5964.1; -.
DR PIR: A61196; A61196.
DR PIR: P0246; P0246.
DR PIR: P0329; P0329.
DR HSSP: P27958; 1HE1.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019028; C:viral capsid; IEA.
DR GO: GO:0019031; C:viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferrase activity; IEA.
DR GO: GO:0006118; F:electron transport; IEA.
DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; F:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_Ser_tryptsin.
DR InterPro: IPR000345; Cys_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR00745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS5a.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; Peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_Ps.
DR InterPro: IPR007094; RNA_pol_PsVlr.
DR Pfam: PF01543; HCV_capsid.1.
DR Pfam: PF01542; HCV_core.1.
DR Pfam: PF01539; HCV_env.1.
DR Pfam: PF01560; HCV_NS1.1.
DR Pfam: PF01538; HCV_NS2.1.
DR Pfam: PF02907; HCV_NS3.1.
DR Pfam: PF01006; HCV_NS4a.1.
DR Pfam: PF01001; HCV_NS5a.1.
DR Pfam: PF01506; HCV_NS5a.1.

```

DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; Viral\_RBP; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C9DB93B60C7 CRC64;

Query Match 96.7%; Score 1537; DB 12; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 3.1e-126;  
 Matches 288; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGITKVFYFRAQGLIRACMLVRKAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 60  
 DB 904 AGITRMYPFVRAQGLIRACMLVRKAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 963  
 QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 120  
 DB 964 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 1023  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCMVTFH 180  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCMVTFH 1083  
 QY 181 GAGSKTLAGPKPITQMTYTNVDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVR 240  
 DB 1084 GAGSKTLAGPKPITQMTYTNVDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVR 1143  
 QY 241 RRGDSRGSLSPRPVSYLKSGSGPFLCPSGHAGVIFRAAVCTRGVAKAVDPIVESHMET 300  
 DB 1144 RRGDSRGSLSPRPVSYLKSGSGPFLCPSGHAGVIFRAAVCTRGVAKAVDPIVESHMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 11  
 ID 070815 PRELIMINARY; PRT; 361 AA.

AC 070815;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DE Polypeptide (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCB1\_Taxid=11103;  
 RN NCBI [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98321154; PubMed=9656998;  
 RA Yamada K., Mori A., Seki M., Kimura J., Yuasa S., Matsunura Y.,  
 RA Miyamura T.;  
 RT "Critical point mutations for hepatitis C virus NS3 proteinase";  
 RL Virology 246:104-112(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,  
 RA Miyamura T.;  
 RT "Enzymatic characterization of purified NS3 serine proteinase of  
 RT hepatitis C virus expressed in Escherichia coli";  
 RL FEBS Lett. 378:37-42(1998).  
 DR EMBL; AB013620; BAA28498.1; -.  
 DR HSP; P27958; IHEI.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_crypstin.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.  
 FT NON\_TER 1  
 FT NON\_TER 361  
 SQ SEQUENCE 361 AA; 38336 MW; 87DC310C76F4BCC3 CRC64;

Query Match 96.5%; Score 1534; DB 12; Length 361;  
 Best Local Similarity 95.0%; Pred. No. 3.5e-127;  
 Matches 288; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVFYFRAQGLIRACMLVRKAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 60  
 DB 5 AGITRMYPFVRAQGLIRACMLVRKAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 64  
 QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 120  
 DB 65 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 124  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCMVTFH 180  
 DB 125 RLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCMVTFH 184  
 QY 181 GAGSKTLAGPKPITQMTYTNVDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVR 240  
 DB 185 GAGSKTLAGPKPITQMTYTNVDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVR 244  
 QY 241 RRGDSRGSLSPRPVSYLKSGSGPFLCPSGHAGVIFRAAVCTRGVAKAVDPIVESHMET 300  
 DB 245 RRGDSRGSLSPRPVSYLKSGSGPFLCPSGHAGVIFRAAVCTRGVAKAVDPIVESHMET 304  
 QY 301 TMR 303  
 DB 305 TMR 307

RESULT 12  
 ID 093F4 PRELIMINARY; PRT; 3008 AA.

AC 093F4;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE Polypeptide (Fragment).  
 DE Genome polypeptide.  
 GN MD34.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCB1\_Taxid=11103;  
 RN NCBI [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD34;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.,  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 RT progression";  
 RL Submitted (Nov-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC EMBL; AF208024; AAF61205.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; F80329; F80329.  
 DR HSP; P26663; IUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008025; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR003045; Cys\_ser\_trypsin.  
 DR InterPro: IPR00345; CysC\_heme\_B5.  
 DR InterPro: IPR00410; DEAD.  
 DR InterPro: IPR002523; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 DR Hydroxylase; Nonstructural protein; Transferase; Transmembrane.  
 DR RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 DR KW  
 DR SEQUENCE 3008 AA; 326834 MW; 99AE05E14C3109F4 CRC64;  
 SO  
 Query Match 96.5%; Score 1534; DB 12; Length 3008;  
 Best Local Similarity 95.7%; Pred. No. 5,7e-126;  
 Matches 290; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

ID 09J3H3 PRELIMINARY; FRT; 3010 AA.  
 AC 09J3H3;  
 DT 01-OCT-2000 (TRENBLREL.15, Created)  
 DT 01-OCT-2000 (TRENBLREL.15, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL.25, Last annotation update)  
 DE Genome polypeptide.  
 DE Hepatitis C virus.  
 DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepatitis C.  
 CC NCBI\_TaxID=11103;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MD19.  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 RT Characteristics of hepatitis C viral genome associated with disease  
 RT progression.  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL: AF207760; AAE65950.1; -.  
 DR PIR: A61196; A61196.  
 DR HSRP: P26663; IUXP.  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR003045; Cys\_ser\_trypsin.  
 DR InterPro: IPR00410; DEAD.  
 DR InterPro: IPR002523; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR ATP-binding; Envelope protein; Glycoprotein; Nonstructural protein;  
 DR Coat protein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 DR KW

RESULT 13  
 09J3H3

SEQ SEQUENCE 3010 AA; 327234 MW; 44C3467649C8BD CRC64;  
Query Match 96.5%; Score 1534; DB 12; Length 3010;  
Best Local Similarity 94.4%; Pred. No. 5,7e-126;  
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;  
QY 1 AGITKVPYFRAQGLIRACMLVRKAAGHYVQMAFMALALGTGYVDHLTPLOMAHAG 60  
DB 904 AGITRVYFRAQGLIRACMLVRKAAGHYVQMAFMALALGTGYVDHLTPLOMAHAG 963  
QY 61 LRDLAAVEPVFESDMETKIITWGDPTAACGDIISGLPVASRGREILLGPADNFEQGW 120  
DB 964 LRDLAAVEPVFESDMETKIITWGDPTAACGDIISGLPVASRGREILLGPADNFEQGW 1023  
QY 121 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFLATCVNGVCTVPH 180  
DB 1024 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFLATCVNGVCTVPH 1083  
QY 181 GAGSKTLAGKPGITOMYTNVDDLVGMQAPGARSWTPCTGSSDLYLTRHADYIPVR 240  
DB 1084 GAGAKTLAGKPGITOMYTNVDDLVGMQAPGARSWTPCTGSSDLYLTRHADYIPVR 1143  
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 300  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206  
RESULT 14  
09J3H2 PRELIMINARY; PRT: 3010 AA.  
AC 09J3H2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polypeptide.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MD20;  
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease  
progression.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA (BY SIMILARITY).  
DR EMBL, AF207761, AA65951.1, -.  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.  
DR HSP; P26663; INS3.  
DR GO:0016021; C:integral to membrane; IEA.  
DR GO:0019028; C:viral capsid; IEA.  
DR GO:0019031; C:viral envelope; IEA.  
DR GO:0005524; F:ATP binding; IEA.  
DR GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO:0005489; F:electron transporter activity; IEA.  
DR GO:0003723; F:RNA binding; IEA.  
DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO:0005198; F:structural molecule activity; IEA.  
DR GO:0016740; F:transferase activity; IEA.  
DR GO:0006118; P:electron transport; IEA.  
DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO:0006350; P:transcription; IEA.

DR GO:0019079; P:viral genome replication; IEA.  
DR GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR009003; Cys\_Ser\_lypsin.  
DR InterPro; IPR000345; CysC\_heme\_BS.  
DR InterPro; IPR01410; DEAD\_  
DR InterPro; IPR02522; HCV capsid.  
DR InterPro; IPR002521; HCV capsid.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR002518; HCV NS1.  
DR InterPro; IPR000745; HCV NS4a.  
DR InterPro; IPR001490; HCV NS4b.  
DR InterPro; IPR002668; HCV NS5a.  
DR InterPro; IPR002166; HCV RdRp.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA\_pol\_D5\_P6.  
DR InterPro; IPR007094; RNA\_pol\_P5vir.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV NS1; 1.  
DR Pfam; PF01538; HCV NS2; 1.  
DR Pfam; PF02907; HCV NS3; 1.  
DR Pfam; PF01006; HCV NS4a; 1.  
DR Pfam; PF01001; HCV NS4b; 1.  
DR Pfam; PF01506; HCV NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRp; 1.  
DR ProDom; PD186062; HCV NS1; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polypeptide; RNA-directed RNA polymerase; transferase; Transmembrane.  
SEQ SEQUENCE 3010 AA; 326763 MW; 1A48BE4BE51440D0 CRC64;  
Query Match 96.5%; Score 1533; DB 12; Length 3010;  
Best Local Similarity 95.0%; Pred. No. 7e-126;  
Matches 286; Conservative 9; Mismatches 6; Indels 0; Gaps 0;  
QY 1 AGITKVPYFRAQGLIRACMLVRKAAGHYVQMAFMALALGTGYVDHLTPLOMAHAG 60  
DB 904 AGITRVYFRAQGLIRACMLVRKAAGHYVQMAFMALALGTGYVDHLTPLOMAHAG 963  
QY 61 LRDLAAVEPVFESDMETKIITWGDPTAACGDIISGLPVASRGREILLGPADNFEQGW 120  
DB 964 LRDLAAVEPVFESDMETKIITWGDPTAACGDIISGLPVASRGREILLGPADNFEQGW 1023  
QY 121 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFLATCVNGVCTVPH 180  
DB 1024 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFLATCVNGVCTVPH 1083  
QY 181 GAGSKTLAGKPGITOMYTNVDDLVGMQAPGARSWTPCTGSSDLYLTRHADYIPVR 240  
DB 1084 GAGAKTLAGKPGITOMYTNVDDLVGMQAPGARSWTPCTGSSDLYLTRHADYIPVR 1143  
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 300  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206  
RESULT 15  
09J3H2 PRELIMINARY; PRT: 3010 AA.  
AC 09J3H2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polypeptide.

OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxId=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD12;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBMITTER: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; AF207753; AAF5943.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; PS0329; PS0329.  
 DR HSBP; P2663; IUXP.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:Electron transporter activity; IEA.  
 DR GO; GO:0016787; F:Hydrolase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; F:electron transport; IEA.  
 DR GO; GO:0006508; F:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:Viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys Ser tlypsin.  
 DR InterPro; IPR000345; CysC\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRP.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SMO0487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolyase; Nonstructural protein; Polyprotein;  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326692 MW; 074098DB305AF1A9 CRC64;

Query Match

96.5%; Score 1533; DB 12; Length 3010;

Best Local Similarity 95.4%; Pred. No. 7e-126;  
 Matches 289; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 AGITKVPYFVPAQGLIRACMLVRKAGGHVYQAFKLAALTGTYYDHLTPLODWAHAG 60  
 DB 904 AGITRVYFVPAQGLIRACMLVRKAGGHVYQAFKLAALTGTYYDHLTPLODWAHAG 963  
 QY 61 LRDIAVAEPIFSDMEVKIITWGDTPAACDIIISGLPVASRRREILLGPADNFEQGN 120  
 DB 964 LRDIAVAEPIFSDMEVKIITWGDTPAACDIIISGLPVASRRREILLGPADNFEQGN 1023  
 QY 121 RLAPITAYSOQTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVFH 180  
 DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVFH 1083  
 QY 181 GAGSKTLAGPKPTTQMTNTVDOLVGMQAPPGARSMTPCTCGSSDILYTRHADVTPVR 240  
 DB 1084 GAGSKTLAGPKPTTQMTNTVDOLVGMQAPPGARSMTPCTCGSSDILYTRHADVTPVR 1143  
 QY 241 RRGDSRGSLSPRVSYLKSSGGPILCPSGHAGIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 1144 RRGDSRGSLSPRVSYLKSSGGPILCPSGHAGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 301 TWR 303  
 DB 1204 TWR 1206

Search completed: May 6, 2004, 09:35:45  
 Job time : 29.339 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 11.8923 seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-10  
Perfect score: 1589  
Sequence: 1 AGITKVYFVRAGGLIRAC.....RGVAKAVDFIPVESMETWR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patente AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCtUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1530	96.3	2201	4 US-09-539-601-6	Sequence 6, Appl
2	1530	96.3	2201	4 US-09-539-601-15	Sequence 15, Appl
3	1530	96.3	3010	4 US-09-539-601-3	Sequence 3, Appl
4	1530	96.3	3010	4 US-09-539-601-21	Sequence 21, Appl
5	1530	96.3	3010	4 US-09-539-601-27	Sequence 27, Appl
6	1527	96.1	1692	3 US-09-263-933-4	Sequence 4, Appl
7	1527	96.1	1692	3 US-09-263-933-2	Sequence 2, Appl
8	1527	96.1	2307	3 US-09-263-933-11	Sequence 11, Appl
9	1527	96.1	2307	3 US-09-263-933-12	Sequence 12, Appl
10	1524	95.9	1692	4 US-09-263-933-9	Sequence 9, Appl
11	1524	95.9	2307	3 US-09-263-933-11	Sequence 11, Appl
12	1524	95.9	2307	3 US-09-263-933-9	Sequence 9, Appl
13	1524	95.9	2307	3 US-09-263-933-9	Sequence 9, Appl
14	1523	95.8	3010	4 US-09-539-601-33	Sequence 33, Appl
15	1515	95.3	1692	3 US-09-263-933-18	Sequence 18, Appl
16	1515	95.3	1692	3 US-09-263-933-15	Sequence 16, Appl
17	1515	95.3	2307	4 US-09-263-933-15	Sequence 16, Appl
18	1515	95.3	2307	4 US-09-263-933-15	Sequence 16, Appl
19	1504	94.7	3010	3 US-09-014-816-3	Sequence 3, Appl
20	1478	93.0	2013	1 US-08-324-977-12	Sequence 12, Appl
21	1478	93.0	2013	2 US-08-384-616-12	Sequence 12, Appl
22	1478	93.0	2013	2 US-08-384-616-12	Sequence 12, Appl
23	1478	93.0	2013	3 US-09-315-850-12	Sequence 12, Appl
24	1478	93.0	2201	4 US-08-952-881A-2	Sequence 2, Appl
25	1478	93.0	2620	1 US-08-324-977-32	Sequence 32, Appl
26	1478	93.0	2620	2 US-08-384-616-32	Sequence 32, Appl
27	1478	93.0	2620	2 US-08-904-686A-32	Sequence 32, Appl

28	1478	93.0	2620	3 US-09-315-850-32	Sequence 32, Appl
29	1478	93.0	2621	1 US-08-324-977-36	Sequence 36, Appl
30	1478	93.0	2621	2 US-08-384-616-36	Sequence 36, Appl
31	1478	93.0	2621	2 US-08-904-686A-36	Sequence 36, Appl
32	1478	93.0	2621	2 US-09-315-850-36	Sequence 36, Appl
33	1478	93.0	3010	1 US-08-324-977-2	Sequence 2, Appl
34	1478	93.0	3010	1 US-08-324-977-14	Sequence 14, Appl
35	1478	93.0	3010	2 US-08-384-616-2	Sequence 2, Appl
36	1478	93.0	3010	2 US-08-384-616-14	Sequence 14, Appl
37	1478	93.0	3010	2 US-08-904-686A-2	Sequence 2, Appl
38	1478	93.0	3010	2 US-08-904-686A-14	Sequence 14, Appl
39	1478	93.0	3010	2 US-09-315-850-2	Sequence 2, Appl
40	1478	93.0	3010	3 US-09-315-850-14	Sequence 14, Appl
41	1405	88.4	3012	3 US-08-811-566-2	Sequence 2, Appl
42	1405	88.4	3012	4 US-09-034-756-2	Sequence 2, Appl
43	1402	88.2	1648	1 US-08-188-281B-12	Sequence 12, Appl
44	1402	88.2	1648	1 PCT-US94-07280-12	Sequence 12, Appl
45	1402	88.2	1648	5 PCT-US95-01087-12	Sequence 12, Appl

## ALIGNMENTS

RESULT 1  
US-09-539-601-6  
; Sequence 6, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartenhager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; EARLIER FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-6

Query Match 96.3%; Score 1530; DB 4; Length 2201;  
Best Local Similarity 95.0%; Pred. No. 2.5e-144;  
Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY	1	AGITKVYFVRAGGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYDHLTPLODMAHAG	60
DB	95	AGITKVYFVRAGGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYDHLTPLODMAHAG	154
QY	61	LEDLAAVEPVYFSDMEVKIITWAGDTAACGDIISGLFVSARGSEIILGPADNFEQGW	120
DB	155	LEDLAAVEPVYFSDMEVKIITWAGDTAACGDIISGLFVSARGSEIILGPADNFEQGW	214
QY	121	RLAATITVSOOTRGLIITSLTGRDNQVEGVVVSATOSFLATCVNGVCMTFPH	180
DB	215	RLAATITVSOOTRGLIITSLTGRDNQVEGVVVSATOSFLATCVNGVCMTFPH	274
QY	181	GAGSKTLGPKPIITONTVNDQDVLVQAPPGASMTPTCGSSDLYLVRHADIVFR	240
DB	275	GAGSKTLGPKPIITONTVNDQDVLVQAPPGASMTPTCGSSDLYLVRHADIVFR	334
QY	241	RQGDGRGSLSRPVSYLKGSGGGLCPSGHAGVIFPAVCTRGVAKAVDFIPVESMET	300
DB	335	RQGDGRGSLSRPVSYLKGSGGGLCPSGHAGVIFPAVCTRGVAKAVDFIPVESMET	394
QY	301	TWR 303	
DB	395	TWR 397	

RESULT 2

US-09-539-601-15  
 ; Sequence 15, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 15  
 ; LENGTH: 2201  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-15

Query Match 96.3%; Score 1530; DB 4; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 2,5e-144;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 |||||  
 DB 95 AGITKVPYFVRAHGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 154  
 |||||  
 QY 61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVARSRGREIILGPADNFEQGW 120  
 |||||  
 DB 155 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVARSRGREIILGPADNFEQGW 214  
 |||||  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLATCVNGVCWTVFH 180  
 |||||  
 DB 215 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLATCVNGVCWTVFH 274  
 |||||  
 QY 181 GAGSKTAGPKGPIITQMTNVDDLVGMQAPPGARSPTCTCGSSDLVYTRHADVI PVR 240  
 |||||  
 DB 275 GAGSKTAGPKGPIITQMTNVDDLVGMQAPPGARSPTCTCGSSDLVYTRHADVI PVR 334  
 |||||  
 QY 241 RRGDSRGSILSPRVSYLKSGSGGELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 |||||  
 DB 335 RRGDSRGSILSPRVSYLKSGSGGELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 394  
 |||||  
 QY 301 TMR 303  
 |||||  
 DB 395 TMR 397

RESULT 3  
 US-09-539-601-3  
 ; Sequence 3, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 3010  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-3

Query Match 96.3%; Score 1530; DB 4; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 3,9e-144;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 60

DB 904 AGITKVPYFVRAHGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 |||||  
 QY 61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVARSRGREIILGPADNFEQGW 120  
 |||||  
 DB 964 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVARSRGREIILGPADNFEQGW 1023  
 |||||  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLATCVNGVCWTVFH 180  
 |||||  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLATCVNGVCWTVFH 1083  
 |||||  
 QY 181 GAGSKTAGPKGPIITQMTNVDDLVGMQAPPGARSPTCTCGSSDLVYTRHADVI PVR 240  
 |||||  
 DB 1084 GAGSKTAGPKGPIITQMTNVDDLVGMQAPPGARSPTCTCGSSDLVYTRHADVI PVR 1143  
 |||||  
 QY 241 RRGDSRGSILSPRVSYLKSGSGGELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 |||||  
 DB 1144 RRGDSRGSILSPRVSYLKSGSGGELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 |||||  
 QY 301 TMR 303  
 |||||  
 DB 1204 TMR 1206

RESULT 4  
 US-09-539-601-21  
 ; Sequence 21, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 21  
 ; LENGTH: 3010  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-21

Query Match 96.3%; Score 1530; DB 4; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 3,9e-144;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 |||||  
 DB 904 AGITKVPYFVRAHGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 |||||  
 QY 61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVARSRGREIILGPADNFEQGW 120  
 |||||  
 DB 964 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVARSRGREIILGPADNFEQGW 1023  
 |||||  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLATCVNGVCWTVFH 180  
 |||||  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLATCVNGVCWTVFH 1083  
 |||||  
 QY 181 GAGSKTAGPKGPIITQMTNVDDLVGMQAPPGARSPTCTCGSSDLVYTRHADVI PVR 240  
 |||||  
 DB 1084 GAGSKTAGPKGPIITQMTNVDDLVGMQAPPGARSPTCTCGSSDLVYTRHADVI PVR 1143  
 |||||  
 QY 241 RRGDSRGSILSPRVSYLKSGSGGELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 |||||  
 DB 1144 RRGDSRGSILSPRVSYLKSGSGGELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 |||||  
 QY 301 TMR 303  
 |||||  
 DB 1204 TMR 1206



RESULT 5  
 US-09-539-601-27  
 ; Sequence 27, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartschlagel, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 27  
 ; LENGTH: 3010  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-27

Query Match 96.3%; Score 1530; DB 4; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 3.5e-144;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYVYDHLTPLODMAHAG 60  
 DB 904 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYVYDHLTPLODMAHAG 963  
 QY 61 LRDLAAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 120  
 DB 964 LRDLAAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 1023  
 QY 121 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 180  
 DB 1024 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 1083  
 QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGMQAPPGASMTPTCTGSSDLVLTTRHADVIPVR 240  
 DB 1084 GAGSKTLAGPKPITQMTYTNVDQDLVGMQAPPGASMTPTCTGSSDLVLTTRHADVIPVR 1143  
 QY 241 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHAGVITRAVCTRGVAKANDPIPVESMET 300  
 DB 1144 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHAGVITRAVCTRGVAKANDPIPVESMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 6  
 US-09-263-933-4  
 ; Sequence 4, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/263,933  
 ; CURRENT FILING DATE: 1999-03-08  
 ; EARLIER APPLICATION NUMBER: 09/129,611  
 ; EARLIER FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 US-09-263-933-4

Query Match 96.1%; Score 1527; DB 3; Length 1692;

Best Local Similarity 94.4%; Pred. No. 3.5e-144;  
 Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYVYDHLTPLODMAHAG 60  
 DB 183 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYVYDHLTPLODMAHAG 242  
 QY 61 LRDLAAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 120  
 DB 243 LRDLAAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
 QY 121 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 180  
 DB 303 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 362  
 QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGMQAPPGASMTPTCTGSSDLVLTTRHADVIPVR 240  
 DB 363 GAGSKTLAGPKPITQMTYTNVDQDLVGMQAPPGASMTPTCTGSSDLVLTTRHADVIPVR 422  
 QY 241 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHAGVITRAVCTRGVAKANDPIPVESMET 300  
 DB 423 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHAGVITRAVCTRGVAKANDPIPVESMET 482  
 QY 301 TMR 303  
 DB 483 TMR 485

RESULT 7  
 US-09-919-901-4  
 ; Sequence 4, Application US/09919901  
 ; Patent No. 6599738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263,933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 US-09-919-901-4

Query Match 96.1%; Score 1527; DB 4; Length 1692;  
 Best Local Similarity 94.4%; Pred. No. 3.5e-144;  
 Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYVYDHLTPLODMAHAG 60  
 DB 183 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYVYDHLTPLODMAHAG 242  
 QY 61 LRDLAAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 120  
 DB 243 LRDLAAVEPVIFSDMEVKITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
 QY 121 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 180  
 DB 303 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 362  
 QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGMQAPPGASMTPTCTGSSDLVLTTRHADVIPVR 240

```

Db      363 GAGSKTLAGKGPITQWYTNVDOLVGMQAPPGARSLLPTCGSSDLVYTRHADVI PVR 422
QY      241 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVGFRAAVCTRGVAKAVDFIPVESMET 300
Db      423 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVGFRAAVCTRGVAKAVDFIPVESMET 482
QY      301 TMR 303
Db      483 TMR 485

```

```

RESULT 8
US-09-263-933-2
; Sequence 2, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-06
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-2

```

```

Query Match      96.1%; Score 1527; DB 3; Length 2307;
Best Local Similarity 94.4%; Pred. No. 5,4e-144;
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

```

```

QY      1 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db      275 AGITRVYFVRAQGLIHACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 334
QY      61 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRRREIILGPADNPEGQM 120
Db      335 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRRREIILGPADNPEGQM 394
QY      121 RLAPITAYSQOTRGILGCIITSLTGDKNOVEGEVQVSTATQSFATCVNGVCMWTVYH 180
Db      395 RLAPITAYSQOTRGILGCIITSLTGDKNOVEGEVQVSTATQSFATCVNGVCMWTVYH 454
QY      181 GAGSKTLAGKGPITQWYTNVDOLVGMQAPPGARSLLPTCGSSDLVYTRHADVI PVR 240
Db      455 GAGSKTLAGKGPITQWYTNVDOLVGMQAPPGARSLLPTCGSSDLVYTRHADVI PVR 514
QY      241 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVGFRAAVCTRGVAKAVDFIPVESMET 300
Db      515 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVGFRAAVCTRGVAKAVDFIPVESMET 574
QY      301 TMR 303
Db      575 TMR 577

```

```

RESULT 9
US-09-919-901-2
; Sequence 2, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

```

```

; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-2

```

```

Query Match      96.1%; Score 1527; DB 4; Length 2307;
Best Local Similarity 94.4%; Pred. No. 5,4e-144;
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

```

```

QY      1 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db      275 AGITRVYFVRAQGLIHACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 334
QY      61 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRRREIILGPADNPEGQM 120
Db      335 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRRREIILGPADNPEGQM 394
QY      121 RLAPITAYSQOTRGILGCIITSLTGDKNOVEGEVQVSTATQSFATCVNGVCMWTVYH 180
Db      395 RLAPITAYSQOTRGILGCIITSLTGDKNOVEGEVQVSTATQSFATCVNGVCMWTVYH 454
QY      181 GAGSKTLAGKGPITQWYTNVDOLVGMQAPPGARSLLPTCGSSDLVYTRHADVI PVR 240
Db      455 GAGSKTLAGKGPITQWYTNVDOLVGMQAPPGARSLLPTCGSSDLVYTRHADVI PVR 514
QY      241 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVGFRAAVCTRGVAKAVDFIPVESMET 300
Db      515 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVGFRAAVCTRGVAKAVDFIPVESMET 574
QY      301 TMR 303
Db      575 TMR 577

```

```

RESULT 10
US-09-263-933-11
; Sequence 11, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-06
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-11

```

```

Query Match      95.9%; Score 1524; DB 3; Length 1692;
Best Local Similarity 94.1%; Pred. No. 6,9e-144;
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

```

Fri May 7 13:36:59 2004

us-10-650-585-10.rai

Page 5

[illegible]

```

RESULT 11
US-09-919-901-11
Sequence 11, Application US/09919901
Patent No. 6599738
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/919,901
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-02-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
US-09-919-901-11

```

	Query Match	Similarity	%	Score	DB	Length	Mismatches	Indels	Gaps
Best Local	285	Conservative	95.9%	1524	1692				
Matches	285	Conservative	94.1%	144	1692				
	12	Mismatches	6	Indels	0	Gaps	0		
Cy	1	AGTGTGPPFVFNAGGIIIRACMLVYRKAGAGHYVOMAFKLAALTGTVYDHLPTGQMAHAG	60						
Db	183	AGTTRPPIFFVRAAGGIIHACMLVYRKAGAGHYVOMAFKALGALITGIYINHLPLVRMAHAG	242						
Cy	61	LRLDAVAEPEVLFSDMEVKIITGWADTAACGDIISGLPVSARGRSEILIGPADNEEGGGM	120						
Db	243	LRLDAVAEPEVLFSDMETKIITGWADTAACGDIISGLPVSARGRSEIILIGPADSLDEGGGM	302						
Cy	121	RLAAPTATVYSOOTRLDLCGIIITSLGRBDNQVEGEVGVSTATQSEFLATCVAGVCMATFEH	180						
Db	303	RLAAPTATVYSOOTRLDLCGIIITSLGRBDNQVEGEVGVSTATQSEFLATCVAGVCMATFEH	362						
Cy	181	GAGSRTLAGPGEPIITOMATVYDQDLVGMQAPGASSMPTCTGSSDLYLVTHADVIFPR	240						
Db	363	GAGSRTLAGPGEPIITOMATVYDQDLVGMQAPGASSLPCTGSSDLYLVTHADVIFPR	422						

Db 423 RRDSDSGSLSPRPVSYLKGSA GGPLCPGSHAVGIFPAAVCTRGVAKAVDFPVESMET 482  
QY 301 TMR 303  
Db 483 TMR 485

```

RESULT 12
US-09-263-933-9
: Sequence 9, Application US/09263933
: Patent No. 6280940
: GENERAL INFORMATION:
: APPLICANT: Potts, Karen E.
: APPLICANT: Jackson, Roberta L.
: APPLICANT: Patlick, Amy K.
: TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
: TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
: FILE REFERENCE: 0125-0005A
: CURRENT APPLICATION NUMBER: US/09/263,933
: CURRENT FILING DATE: 1999-03-08
: EARLIER APPLICATION NUMBER: 09/129,611
: EARLIER FILING DATE: 1998-08-05
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 9
: LENGTH: 2307
: TYPE: PRT
: ORGANISM: Artificial Sequence
US-09-263-933-9

```

```

Query Match      95.9%; Score 1524; DB 3; Length 2307;
Best Local Similarity 94.1%; Pred. No. 1,1e-143;
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY      1 AGITKVPYFVRAOGLIRACMLVKKAGHYVQMAFMKLAALTGTYYVDLTPIDMAHAG 60
Db      275 AGITRVYFVRAOGLIRACMLVKKAGHYVQMAFMKLGALTGTYYVNHLLTPIDMAHAG 334

QY      61 LRDLAAVEEVTISDMEVKIITWGADPAAACGDIISGLPVARRGRELLIGPADNFEQGW 120
Db      335 LRDLAAVEEVEVPSDMEVKIITWGADPAAACGDIISGLPVARRKRELLIGPADSLEGRW 394

QY      121 RLAPAFAYSQQRGLLGCIITSLTGDKQVGEVGVSTATQSPFATCNGVCWTFVH 180
Db      395 RLAPAFAYSQQRGLLGCIITSLTGDKQVGEVGVSTATQSPFATCNGVCWTFVYH 454

QY      181 GAGSKTLAGEKPIITQMTNVDDLVGMQAPPGARSKTPTCCGSSDLYLVTRHADVIPIVR 240
Db      455 GAGSKTLAGEKPIITQMTNVDDLVGMQAPPGARSKTPTCCGSSDLYLVTRHADVIPIVR 514

QY      241 RRGDSRSLSPRPVSLKSGSGGPIICSGGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
Db      515 RRGDSRSLSPRPVSLKSGSAGGPIICSGGHAVGIFRAAVCTRGVAKAVDFIPVESMET 574

QY      301 TMR 303
Db      575 TMR 577

RESULT 13
US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen B.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; PUBLICATION NUMBER: 2009/0000000

```

PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-9

Query Match 95.3%; Score 1524; DB 4; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 1.1e-143;  
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODWAHAG 60  
DB 275 AGITRVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODWAHAG 334  
QY 61 LRDLAVALVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPDNFEQGW 120  
DB 335 LRDLAVALVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPDNFEQGW 394  
QY 121 RLAPITAYSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVYH 180  
DB 395 RLAPITAYSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVYH 454  
QY 181 GAGSKTLAAGPKGPIITQWYTNVDOLVGMQAPPGARSMTPTCGSSDLVYTRHADVI 240  
DB 455 GAGSKTLAAGPKGPIITQWYTNVDOLVGMQAPPGARSMTPTCGSSDLVYTRHADVI 514  
QY 241 RRGDSRGSLLSPRVSYLKSGSGGFLPCPSGHAAGVIFRAAVCTRGVAKAVDPVVESEMT 300  
DB 515 RRGDSRGSLLSPRVSYLKSGSGGFLPCPSGHAAGVIFRAAVCTRGVAKAVDPVVESEMT 574  
QY 301 TMR 303  
DB 575 TMR 577

RESULT 14  
US-09-539-601-33  
Sequence 33, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Barteneschlager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
EARLIER FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 33  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-33

Query Match 95.8%; Score 1523; DB 4; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 2e-143;  
Matches 287; Conservative 8; Mismatches 8; Indels 0; Gaps 0;  
QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODWAHAG 60  
DB 904 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODWAHAG 963  
QY 61 LRDLAVALVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPDNFEQGW 120

DB 964 LRDLAVALVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPDNFEQGW 1023  
QY 121 RLAPITAYSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVYH 180  
DB 1024 RLAPITAYSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVYH 1083  
QY 181 GAGSKTLAAGPKGPIITQWYTNVDOLVGMQAPPGARSMTPTCGSSDLVYTRHADVI 240  
DB 1084 GAGSKTLAAGPKGPIITQWYTNVDOLVGMQAPPGARSMTPTCGSSDLVYTRHADVI 1143  
QY 241 RRGDSRGSLLSPRVSYLKSGSGGFLPCPSGHAAGVIFRAAVCTRGVAKAVDPVVESEMT 300  
DB 1144 RRGDSRGSLLSPRVSYLKSGSGGFLPCPSGHAAGVIFRAAVCTRGVAKAVDPVVESEMT 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 15  
US-09-263-933-18  
Sequence 18, Application US/09263933  
Patent No. 6280340  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Potts, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
EARLIER FILING DATE: 1999-03-08  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 18  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-18

Query Match 95.3%; Score 1515; DB 3; Length 1692;  
Best Local Similarity 93.7%; Pred. No. 5.6e-143;  
Matches 284; Conservative 12; Mismatches 7; Indels 0; Gaps 0;  
QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODWAHAG 60  
DB 183 AGITRVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODWAHAG 242  
QY 61 LRDLAVALVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPDNFEQGW 120  
DB 243 LRDLAVALVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPDNFEQGW 302  
QY 121 RLAPITAYSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVYH 180  
DB 303 RLAPITAYSQOTRGLIGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVYH 362  
QY 181 GAGSKTLAAGPKGPIITQWYTNVDOLVGMQAPPGARSMTPTCGSSDLVYTRHADVI 240  
DB 363 GAGSKTLAAGPKGPIITQWYTNVDOLVGMQAPPGARSMTPTCGSSDLVYTRHADVI 422  
QY 241 RRGDSRGSLLSPRVSYLKSGSGGFLPCPSGHAAGVIFRAAVCTRGVAKAVDPVVESEMT 300  
DB 423 RRGDSRGSLLSPRVSYLKSGSGGFLPCPSGHAAGVIFRAAVCTRGVAKAVDPVVESEMT 482  
QY 301 TMR 303  
DB 483 TMR 485

Search completed: May 6, 2004, 09:39:01  
Job time: 12.8923 secs

Fri May 7 13:36:59 2004

us-10-650-585-10.rai

---

Page 7



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 : Search time 30.9958 Seconds  
(without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-10  
Perfect score: 1589  
Sequence: 1 AGITKPYVRAQGLIRACM.....RGVAKAVDFIPVSMETWR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCF\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCFUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1589	100.0	303	US-10-017-736-10 Sequence 10, Appl
2	1589	100.0	303	US-10-650-585-10 Sequence 16, Appl
3	1589	100.0	303	US-10-017-736-10 Sequence 4, Appl
4	1589	100.0	303	US-10-650-585-10 Sequence 17, Appl
5	1589	100.0	303	US-10-017-736-10 Sequence 14, Appl
6	1589	100.0	303	US-10-650-585-10 Sequence 14, Appl
7	1589	100.0	303	US-10-017-736-10 Sequence 13, Appl
8	1589	100.0	303	US-10-650-585-10 Sequence 13, Appl
9	1589	100.0	303	US-10-017-736-10 Sequence 12, Appl
10	1589	100.0	303	US-10-650-585-10 Sequence 12, Appl
11	1589	100.0	303	US-10-017-736-10 Sequence 11, Appl
12	1589	100.0	303	US-10-650-585-10 Sequence 11, Appl
13	1589	100.0	303	US-10-017-736-10 Sequence 2, Appl
14	1589	100.0	303	US-10-650-585-10 Sequence 2, Appl
15	1580	99.4	303	US-10-017-736-10 Sequence 18, Appl

16	1580	99.4	303	US-10-650-585-10 Sequence 18, Appl
17	1579	99.4	303	US-10-017-736-10 Sequence 16, Appl
18	1579	99.4	303	US-10-650-585-10 Sequence 16, Appl
19	1570	98.8	301	US-10-017-736-10 Sequence 17, Appl
20	1570	98.8	301	US-10-650-585-10 Sequence 17, Appl
21	1532	96.4	292	US-10-017-736-10 Sequence 15, Appl
22	1532	96.4	292	US-10-650-585-10 Sequence 15, Appl
23	1530	96.3	2201	US-10-029-907-3 Sequence 3, Appl
24	1530	96.3	2201	US-10-309-561-3 Sequence 3, Appl
25	1530	96.3	3010	US-10-467-000-1 Sequence 1, Appl
26	1527	96.1	1692	US-10-191-866-4 Sequence 4, Appl
27	1527	96.1	1692	US-10-191-866-4 Sequence 4, Appl
28	1527	96.1	2307	US-09-919-901-2 Sequence 2, Appl
29	1527	96.1	2307	US-10-191-866-2 Sequence 2, Appl
30	1524	95.9	1692	US-09-919-901-11 Sequence 11, Appl
31	1524	95.9	1692	US-10-191-866-11 Sequence 11, Appl
32	1524	95.9	2307	US-09-919-901-9 Sequence 9, Appl
33	1524	95.9	2307	US-10-191-866-9 Sequence 9, Appl
34	1515	95.3	1692	US-09-919-901-18 Sequence 18, Appl
35	1515	95.3	1692	US-10-191-866-18 Sequence 18, Appl
36	1515	95.3	2307	US-09-919-901-16 Sequence 16, Appl
37	1515	95.3	2307	US-10-191-866-16 Sequence 16, Appl
38	1478	93.0	2201	US-10-085-476-2 Sequence 2, Appl
39	1405	88.4	3011	US-09-742-659-4 Sequence 4, Appl
40	1405	88.4	3011	US-09-891-894-3 Sequence 3, Appl
41	1405	88.4	3011	US-10-184-150-3 Sequence 3, Appl
42	1405	88.4	3011	US-10-328-997-3 Sequence 3, Appl
43	1405	88.4	3012	US-09-238-076-2 Sequence 2, Appl
44	1405	88.4	3012	US-09-995-937-2 Sequence 2, Appl
45	1405	88.4	3012	US-09-917-563-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-10-017-736-10  
; Sequence 10, Application US/10017736  
; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 10  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-017-736-10

Query Match 100.0%; Score 1589; DB 13; Length 303;  
Best Local Similarity 100.0%; Pred. No. 1.5e+154;  
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AGITKPYVRAQGLIRACM	US-10-017-736-10	Sequence 10, Appl
DB	1	AGITKPYVRAQGLIRACM	US-10-650-585-10	Sequence 16, Appl
QY	61	LRDLAAYVPEVFSDMVEYKIIITGADTAACGDIISGLPVSAARGSEIILGPDNFEQCGM	US-10-017-736-10	Sequence 14, Appl
DB	61	LRDLAAYVPEVFSDMVEYKIIITGADTAACGDIISGLPVSAARGSEIILGPDNFEQCGM	US-10-650-585-10	Sequence 14, Appl
QY	121	RLIAPITAVSQOTRGLIGCIITSLGRDNQVEGEVQVSTATQSPFLATCVNGVCTVYFH	US-10-017-736-10	Sequence 11, Appl
DB	121	RLIAPITAVSQOTRGLIGCIITSLGRDNQVEGEVQVSTATQSPFLATCVNGVCTVYFH	US-10-650-585-10	Sequence 11, Appl
QY	181	GAGSKTLAPKPKPIITQMTNNVQDLYGMAQPAQARSMTCTGSSDLVYVTHADVIPR	US-10-017-736-10	Sequence 18, Appl



Db 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVYTRHADVIPIVR 240  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 QY 301 TMR 303  
 Db 301 TMR 303

## RESULT 2

US-10-650-585-10  
 ; Sequence 10, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 10  
 ; LENGTH: 303  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-10

Query Match 100.0%; Score 1589; DB 16; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 QY 61 LRDLAIVEEPIFSDMEVKIITWGADTPAACGDIISGLPVSARSGREILLGPADNFEQGW 120  
 Db 61 LRDLAIVEEPIFSDMEVKIITWGADTPAACGDIISGLPVSARSGREILLGPADNFEQGW 120  
 QY 121 RLAPITAYSQOTRGILGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
 Db 121 RLAPITAYSQOTRGILGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
 QY 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVYTRHADVIPIVR 240  
 Db 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVYTRHADVIPIVR 240  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 QY 301 TMR 303  
 Db 301 TMR 303

## RESULT 3

US-10-017-736-4  
 ; Sequence 4, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 334  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-017-736-4

Query Match 100.0%; Score 1589; DB 13; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 16 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 QY 61 LRDLAIVEEPIFSDMEVKIITWGADTPAACGDIISGLPVSARSGREILLGPADNFEQGW 120  
 Db 76 LRDLAIVEEPIFSDMEVKIITWGADTPAACGDIISGLPVSARSGREILLGPADNFEQGW 135  
 QY 121 RLAPITAYSQOTRGILGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
 Db 136 RLAPITAYSQOTRGILGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 195  
 QY 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVYTRHADVIPIVR 240  
 Db 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVYTRHADVIPIVR 255  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 QY 301 TMR 303  
 Db 316 TMR 318

## RESULT 4

US-10-650-585-4  
 ; Sequence 4, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 334  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-4

Query Match 100.0%; Score 1589; DB 16; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60.  
 Db 16 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 QY 61 LRDLAIVEEPIFSDMEVKIITWGADTPAACGDIISGLPVSARSGREILLGPADNFEQGW 120  
 Db 76 LRDLAIVEEPIFSDMEVKIITWGADTPAACGDIISGLPVSARSGREILLGPADNFEQGW 135  
 QY 121 RLAPITAYSQOTRGILGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180

```

Db 136 RLAPITAVSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFPLATCVNGVCTVPH 195
QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMPTCTCGSSDLVYTRHADVIPIVR 240
Db 196 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMPTCTCGSSDLVYTRHADVIPIVR 255
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 300
Db 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 315
QY 301 TMR 303
Db 316 TMR 318

```

```

RESULT 5
US-10-017-736-14
; Sequence 14, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-14

```

```

Query Match 100.0%; Score 1589; DB 13; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.7e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db 39 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 98
QY 61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNFEQGM 120
Db 99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNFEQGM 158
QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFPLATCVNGVCTVPH 180
Db 159 RLAPITAVSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFPLATCVNGVCTVPH 218
QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMPTCTCGSSDLVYTRHADVIPIVR 240
Db 219 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMPTCTCGSSDLVYTRHADVIPIVR 278
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 300
Db 279 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 358
QY 301 TMR 303
Db 339 TMR 341

```

```

RESULT 6
US-10-650-585-14
; Sequence 14, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082

```

```

; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-14

```

```

Query Match 100.0%; Score 1589; DB 16; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.7e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db 39 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 98
QY 61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNFEQGM 120
Db 99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNFEQGM 158
QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFPLATCVNGVCTVPH 180
Db 159 RLAPITAVSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFPLATCVNGVCTVPH 218
QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMPTCTCGSSDLVYTRHADVIPIVR 240
Db 219 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMPTCTCGSSDLVYTRHADVIPIVR 278
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 300
Db 279 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 338
QY 301 TMR 303
Db 339 TMR 341

```

```

RESULT 7
US-10-017-736-13
; Sequence 13, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 352
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-13

```

Db 110 LRDLAAVEEVIITSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 169  
 QY 121 RLAPITAYVSOQTRGLIGCIITSLTGRDKQVEGEVQVNSTAQSLATCVCNVCMTVFH 180  
 Db 170 RLAPITAYVSOQTRGLIGCIITSLTGRDKQVEGEVQVNSTAQSLATCVCNVCMTVFH 229  
 QY 181 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIVR 240  
 Db 230 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIVR 289  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 290 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 349  
 QY 301 TMR 303  
 Db 350 TMR 352

## RESULT 8

US-10-650-585-13  
 ; Sequence 13, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-13

## Query Match

Best Local Similarity 100.0%; Score 1589; DB 16; Length 352;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRAKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 50 AGITKVPYFVRAOGLIRACMLVRAKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 109  
 QY 61 LRDLAAVEEVIITSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 Db 110 LRDLAAVEEVIITSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 169  
 QY 121 RLAPITAYVSOQTRGLIGCIITSLTGRDKQVEGEVQVNSTAQSLATCVCNVCMTVFH 180  
 Db 170 RLAPITAYVSOQTRGLIGCIITSLTGRDKQVEGEVQVNSTAQSLATCVCNVCMTVFH 229  
 QY 181 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIVR 240  
 Db 230 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIVR 289  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 290 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 349  
 QY 301 TMR 303  
 Db 350 TMR 352

## RESULT 9

US-10-017-736-12  
 ; Sequence 12, Application US/10017736

; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-017-736-12

Query Match 100.0%; Score 1589; DB 13; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 2e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRAKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 78 AGITKVPYFVRAOGLIRACMLVRAKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 137  
 QY 61 LRDLAAVEEVIITSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 Db 138 LRDLAAVEEVIITSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 197  
 QY 121 RLAPITAYVSOQTRGLIGCIITSLTGRDKQVEGEVQVNSTAQSLATCVCNVCMTVFH 180  
 Db 198 RLAPITAYVSOQTRGLIGCIITSLTGRDKQVEGEVQVNSTAQSLATCVCNVCMTVFH 257  
 QY 181 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIVR 240  
 Db 258 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIVR 317  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 318 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 377  
 QY 301 TMR 303  
 Db 378 TMR 380

## RESULT 10

US-10-650-585-12  
 ; Sequence 12, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-12

Query Match 100.0%; Score 1589; DB 16; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 2e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRAKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60

```

Db      ||||| 78 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 137
Qy      ||||| 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGM 120
Db      ||||| 138 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGM 197
Qy      ||||| 121 RLAPITAYSOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVFFH 180
Db      ||||| 198 RLAPITAYSOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVFFH 257
Qy      ||||| 181 GAGSKTLAAGPKGPIITQMTYNVDQDLVGWQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 240
Db      ||||| 258 GAGSKTLAAGPKGPIITQMTYNVDQDLVGWQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 317
Qy      ||||| 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
Db      ||||| 318 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 377
Qy      ||||| 301 TMR 303
Db      ||||| 378 TMR 380

```

# RESULT 11

```

; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

```

Query Match 100.0%; Score 1589; DB 13; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      ||||| 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db      ||||| 91 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 150
Qy      ||||| 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGM 120
Db      ||||| 151 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGM 210
Qy      ||||| 121 RLAPITAYSOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVFFH 180
Db      ||||| 211 RLAPITAYSOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVFFH 270
Qy      ||||| 181 GAGSKTLAAGPKGPIITQMTYNVDQDLVGWQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 240
Db      ||||| 271 GAGSKTLAAGPKGPIITQMTYNVDQDLVGWQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 330
Qy      ||||| 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
Db      ||||| 331 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 390
Qy      ||||| 301 TMR 303
Db      ||||| 391 TMR 393

```

# RESULT 12

```

US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US2004007066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

```

Query Match 100.0%; Score 1589; DB 16; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      ||||| 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db      ||||| 91 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 150
Qy      ||||| 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGM 120
Db      ||||| 151 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGM 210
Qy      ||||| 121 RLAPITAYSOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVFFH 180
Db      ||||| 211 RLAPITAYSOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVFFH 270
Qy      ||||| 181 GAGSKTLAAGPKGPIITQMTYNVDQDLVGWQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 240
Db      ||||| 271 GAGSKTLAAGPKGPIITQMTYNVDQDLVGWQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 330
Qy      ||||| 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
Db      ||||| 331 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 390
Qy      ||||| 301 TMR 303
Db      ||||| 391 TMR 393

```

# RESULT 13

```

US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2

```

Query Match 100.0%; Score 1589; DB 13; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-154;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
DB 95 AGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 154

QY 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 120  
DB 155 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 214

QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVPH 180  
DB 215 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVPH 274

QY 181 GAGSKTLAGPKGPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 240  
DB 275 GAGSKTLAGPKGPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 334

QY 241 RRGDSRGSLSRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVESMET 300  
DB 335 RRGDSRGSLSRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVESMET 394

QY 301 TMR 303  
DB 395 TMR 397

RESULT 14  
US-10-650-585-2  
; Sequence 2, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-650-585-2

Query Match 100.0%; Score 1589; DB 16; Length 409;  
Best Local Similarity 100.0%; Pred. No. 2,3e-154;  
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
DB 95 AGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 154

QY 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 120  
DB 155 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 214

QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVPH 180  
DB 215 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVPH 274

QY 181 GAGSKTLAGPKGPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 240  
DB 275 GAGSKTLAGPKGPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 334

QY 241 RRGDSRGSLSRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVESMET 300  
DB 335 RRGDSRGSLSRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVESMET 394

QY 301 TMR 303  
DB 395 TMR 397

RESULT 15  
US-10-017-736-18  
; Sequence 18, Application US/10017736  
; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-017-736-18

Query Match 99.4%; Score 1580; DB 13; Length 303;  
Best Local Similarity 99.7%; Pred. No. 1.2e-153;  
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
DB 1 AGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60

QY 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 120  
DB 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARRREIILGPADNFEQGM 120

QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVPH 180  
DB 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCTVPH 180

QY 181 GAGSKTLAGPKGPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 240  
DB 181 GAGSKTLAGPKGPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 240

QY 241 RRGDSRGSLSRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVESMET 300  
DB 241 RRGDSRGSLSRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVESMET 300

QY 301 TMR 303  
DB 301 TMR 303

Search completed: May 6, 2004, 09:43:18  
Job time : 30.9958 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 53.4939 Seconds

(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MAASCGAVFGLALTLSP.....RGVAKAVDFIPVSMETMR 393

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_290Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2053	100.0	393	ABG32184	Abg32184 HCV prote
2	2053	100.0	409	ABG32181	Abg32181 HCV prote
3	1987	96.8	380	ABG32185	Abg32185 HCV prote
4	1967	95.8	3010	AAR82694	Aar82694 Partial H
5	1962	95.6	3010	AAR82622	Aar82622 HCV prote
6	1961	95.5	3010	AAR68864	Aar68864 Hepatitis
7	1951	95.0	2201	ABG30591	Abg30591 Hepatitis
8	1951	95.0	2201	ABG30591	Abg30591 Hepatitis
9	1951	95.0	2201	ABG30600	Abg30600 Hepatitis
10	1951	95.0	2201	ABG30581	Abg30581 Hepatitis
11	1951	95.0	2201	ABG30593	Abg30593 Hepatitis
12	1951	95.0	2201	ABG30582	Abg30582 Hepatitis
13	1951	95.0	2201	ABG30580	Abg30580 Hepatitis
14	1951	95.0	2201	ABG30587	Abg30587 Hepatitis
15	1951	95.0	2201	ABG30599	Abg30599 Hepatitis
16	1951	95.0	2201	ABG30594	Abg30594 Hepatitis
17	1951	95.0	2201	ABG30598	Abg30598 Hepatitis
18	1951	95.0	2201	ABG30595	Abg30595 Hepatitis
19	1951	95.0	3010	ABG32458	Abg32458 Hepatitis
20	1951	95.0	3010	ABG32459	Abg32459 Hepatitis
21	1951	95.0	3010	ABG32451	Abg32451 Hepatitis
22	1951	95.0	3010	ABG32455	Abg32455 Hepatitis
23	1951	95.0	3010	ABG32457	Abg32457 Hepatitis
24	1951	95.0	3010	ABG32460	Abg32460 Hepatitis
25	1951	95.0	3010	ABG32461	Abg32461 Hepatitis

26	1951	95.0	3010	5	ABG32454	Abg32454 Hepatitis
27	1951	95.0	3011	5	ABG32456	Abg32456 Hepatitis
28	1948	94.9	2201	5	ABG30586	Abg30586 Hepatitis
29	1948	94.9	2201	5	ABG30589	Abg30589 Hepatitis
30	1948	94.9	2201	5	ABG30583	Abg30583 Hepatitis
31	1948	94.9	2201	5	ABG30588	Abg30588 Hepatitis
32	1947	94.8	2201	5	ABG30590	Abg30590 Hepatitis
33	1946	94.8	2307	3	AAV70064	AAV70064 Recombina
34	1945	94.7	3010	5	ABG32452	Abg32452 Hepatitis
35	1944	94.7	2201	5	ABG30584	Abg30584 Hepatitis
36	1944	94.7	2201	5	ABG30602	Abg30602 Hepatitis
37	1944	94.7	3010	5	ABG32453	Abg32453 Hepatitis
38	1943	94.6	2307	3	AAV70065	AAV70065 Recombina
39	1943	94.6	3014	2	AAR54099	Aar54099 NANBHV E1
40	1940	94.5	2201	5	ABG30585	Abg30585 Hepatitis
41	1938.5	94.4	768	2	AAR40223	Aar40223 Recombina
42	1938	94.4	3014	2	AAR35207	Aar35207 Recombina
43	1937	94.3	3090	7	ADD67962	ADD67962 EMCV infe
44	1934	94.2	2307	3	AAV70066	AAV70066 Recombina
45	1929	94.0	3010	5	AAR20477	Aar20477 HCV-S1 fu

## ALIGNMENTS

RESULT 1  
ID ABG32184 standard; protein; 393 AA.  
XX AC ABG32184;  
XX AC  
XX 05-NOV-2002 (first entry)  
XX DT  
XX HCV protease NS2/3 truncation mutant 815-1206.  
XX DE  
XX HCV, enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
XX KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
XX KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDMO;  
XX KM chaotropic agent; mutant; mutain.  
XX KM  
XX Hepatitis C virus.  
XX OS Synthetic.  
XX PN  
XX WO200248375-A2.  
XX PD 20-JUN-2002.  
XX PF 13-DEC-2001; 2001WO-CA001796.  
XX PR 15-DEC-2000; 2000US-0256031P.  
XX PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX DR  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
XX PT useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
XX PT  
XX Claim 41; Page 59-60; 67pp; English.  
XX PS  
XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
XX its truncation or a mutated sequence, where the protease is in a solution  
XX comprising a sufficient concentration of lauryldiethylamine oxide (LDMO)  
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX SQ Sequence 393 AA;

Query Match 100.0%; Score 2053; DB 5; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 7, 1e-191;  
 Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAASCGAVFTGLALTLSPYKYLARLIMWLYLITRVAHLQWIPPLNVRGGDAI 60  
 DB 1 MAASCGAVFTGLALTLSPYKYLARLIMWLYLITRVAHLQWIPPLNVRGGDAI 60  
 QY 61 ILITCAHPELIDITLLAIFGLPMLVQAGITKPYEFAAGILRACMLYKAGAHY 120  
 DB 61 ILITCAHPELIDITLLAIFGLPMLVQAGITKPYEFAAGILRACMLYKAGAHY 120  
 QY 121 VQVAFMLAALTGYVVDHLTPLODMNAHGLRDLAVAVEPVPISDEVKITIGADTAA 180  
 DB 121 VQVAFMLAALTGYVVDHLTPLODMNAHGLRDLAVAVEPVPISDEVKITIGADTAA 180  
 QY 181 GDIISGIPVARGRELLGPADNFEQGRLLAPTAAYQQRGILGCIITSLTRDKN 240  
 DB 181 GDIISGIPVARGRELLGPADNFEQGRLLAPTAAYQQRGILGCIITSLTRDKN 240  
 QY 241 QVGEVQVSTATQSFATCVNGVCTVFFHAGSKTLAGKGPITQMTYVNDQDLVQMA 300  
 DB 241 QVGEVQVSTATQSFATCVNGVCTVFFHAGSKTLAGKGPITQMTYVNDQDLVQMA 300  
 QY 301 PPGARSTPCTCGSSDLYLTVTRADVLPVRRRGSGLSPRPVSLKSSGGPLCP 360  
 DB 301 PPGARSTPCTCGSSDLYLTVTRADVLPVRRRGSGLSPRPVSLKSSGGPLCP 360  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
 DB 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

RESULT 2  
 ID ABG32181 standard; protein; 409 AA.

XX AC ABG32181;

XX DT 05-NOV-2002 (first entry)

XX DE HCV protease NS2/3 (810-1206).

XX HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
 KW Chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KW hepatocellular carcinoma; hepatitis; hepatitis B virus infection;  
 KW hepatocellular carcinoma; hepatitis; hepatitis B virus infection;  
 KW chaotropic agent; mutant; mutagen.

XX OS Hepatitis C virus.  
 XX OS Synthetic.  
 XX FH Key Location/Qualifiers  
 FT Peptide 398..409  
 FT /note="Streptavidin tag"  
 XX MO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX WPI; 2002-599511/64.  
 XX DR N-PSDB; ABR90406.  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 42; Fig 1B; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-  
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 residue amino acid 810 to 906, or having a minimal amino acid sequence  
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 NS2/3 protease. Also included are (1) a composition (C) comprising an  
 isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 its truncation or a mutated sequence, where the protease is in a solution  
 comprising a sufficient concentration of lauryldimethylamine oxide (LDAO)  
 to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 protease, involving isolating the protease in the presence of a  
 chaotropic agent, refolding the isolated protease by contacting it with a  
 reducing agent, and LDAO in the presence of reduced concentration of the  
 chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 protease, involving diluting refolded inactive NS2/3 protease in a medium  
 containing an activation detergent to induce auto-cleavage of the NS2/3  
 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 protease, involving incubating the active NS2/3 protease produced by M2  
 for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 cleavage products or their fragments, and measuring the presence or  
 absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 active NS2/3 protease, involving carrying out M3 in the presence of, or  
 absence of the potential inhibitor, comparing the amount of uncleaved  
 NS2/3 protease, cleavage products or their fragments. The protease is  
 useful for detailed biochemical characterisation of the enzymes and in  
 the development of in vitro assays for screening novel inhibitors of  
 NS2/3 protease which are useful as therapeutic agents against HCV  
 infection (which causes chronic liver disease, cirrhosis and end-stage  
 liver disease. M1 is useful for high level production of protease. The  
 present sequence represents the NS2/3 (810-1206) protein, which has a C-  
 terminal streptavidin tag

XX SQ Sequence 409 AA;

Query Match 100.0%; Score 2053; DB 5; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 7, 6e-191;  
 Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAASCGAVFTGLALTLSPYKYLARLIMWLYLITRVAHLQWIPPLNVRGGDAI 60  
 DB 5 MAASCGAVFTGLALTLSPYKYLARLIMWLYLITRVAHLQWIPPLNVRGGDAI 64



QY 61 ILITCAVHPELIDITKLLAIFGPMVLQAGITKVPYVRAQGLIRACMLVKAAGHY 120  
 DB 65 ILITCAVHPELIDITKLLAIFGPMVLQAGITKVPYVRAQGLIRACMLVKAAGHY 124  
 QY 121 VQMAFMKLAALITTYVDHLPQDMAHGLRLAVALVEVIFSDMEVKITWGAADTAAC 180  
 DB 125 VQMAFMKLAALITTYVDHLPQDMAHGLRLAVALVEVIFSDMEVKITWGAADTAAC 184  
 QY 181 GDIIISGIPVARSRRREILGPADEFGQGWRLAPITAVSQOTRGILGCIITSLTRDKN 240  
 DB 185 GDIIISGIPVARSRRREILGPADEFGQGWRLAPITAVSQOTRGILGCIITSLTRDKN 244  
 QY 241 QVGEVQVYSTANOSFLATCVNGVCTVFGASSKTLGKGIITOMYTNVDDLVGMOA 300  
 DB 245 QVGEVQVYSTANOSFLATCVNGVCTVFGASSKTLGKGIITOMYTNVDDLVGMOA 304  
 QY 301 PPGARSMTPCTCGSSDLYLVTTRADYTPVRRGDSRGLSPRPVSYLKSSGGPILCS 360  
 DB 305 PPGARSMTPCTCGSSDLYLVTTRADYTPVRRGDSRGLSPRPVSYLKSSGGPILCS 364  
 QY 361 GHAIVGFRAAVCTRGVAKAVDFIPVSEMTTMR 393  
 DB 365 GHAIVGFRAAVCTRGVAKAVDFIPVSEMTTMR 397

RESULT 3  
 ABG32185  
 ID ABG32185 standard; protein; 380 AA.  
 XX  
 AC ABG32185;  
 XX

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 827-1206.

KM HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethyamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutagen.

OS Hepatitis C virus.  
 XX Synthetic.  
 XX

PN WC0200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WC-CA001796.

PR 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PT Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

DR WPI; 2002-559511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.

PS Claim 41; Page 60-61; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethyamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease, (2) a NS2/3 inhibitory peptide

CC appearing as ABG32185; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 827-1206  
 CC (numbered relative to the full length NS2/3 protein)

SO Sequence 380 AA;

Query Match 96.8%; Score 1987; DB 5; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-184;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ALTLSPYKVLARLIMWLOYLITRVAHLOVPIPNVAGGRALITLCAVHPELIF 73  
 DB 1 ALTLSPYKVLARLIMWLOYLITRVAHLOVPIPNVAGGRALITLCAVHPELIF 60  
 QY 74 DITKLLAIFGPMVLQAGITKVPYVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 133  
 DB 6 DITKLLAIFGPMVLQAGITKVPYVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 120  
 QY 134 TTYVDHLPQDMAHGLRLAVALVEVIFSDMEVKITWGAADTAACDIIISGLPVARR 193  
 DB 121 TTYVDHLPQDMAHGLRLAVALVEVIFSDMEVKITWGAADTAACDIIISGLPVARR 180  
 QY 194 GREIILGPADNEGQWRLAPITAVSQOTRGILGCIITSLTRDKNVEGEVYSTAT 253  
 DB 181 GREIILGPADNEGQWRLAPITAVSQOTRGILGCIITSLTRDKNVEGEVYSTAT 240  
 QY 254 QSFLLATCVNGVCTVFGAGSKTLGAPKGIITOMYTNVDDLVGMOAPPGARSMTPCTCG 313  
 DB 241 QSFLLATCVNGVCTVFGAGSKTLGAPKGIITOMYTNVDDLVGMOAPPGARSMTPCTCG 300  
 QY 314 SSDLYLVTTRADYTPVRRGDSRGLSPRPVSYLKSSGGPILCSGHAIVGFRAAVCT 373  
 DB 301 SSDLYLVTTRADYTPVRRGDSRGLSPRPVSYLKSSGGPILCSGHAIVGFRAAVCT 360  
 QY 374 RGVAKAVDFIPVSEMTTMR 393  
 DB 361 RGVAKAVDFIPVSEMTTMR 380

RESULT 4  
 AAR82694  
 ID AAR82694 standard; protein; 3010 AA.  
 XX  
 AC AAR82694;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 14-NOV-1996 (first entry)  
 XX  
 DE Partial HCV non-structural polypeptide.  
 XX  
 KM proteinase; hepatitis C virus; screening; inhibitor; proteolytic;  
 XX identification; cleavage.

OS	Hepatitis C virus; Virus.	Location/Qualifiers
XX	Key	898..1233
XX	Protein	/note="partial proteinase; see AAF82692"
XX	Protein	922..1907
XX	Protein	/note="partial proteinase; see AAF82693"
XX	JP07184648-A.	
XX	25-JUL-1995.	
XX	05-FEB-1993;	93JP-00018854.
XX	07-FEB-1992;	92JP-00022657.
XX	18-SEP-1992;	92JP-00249240.
XX	04-DEC-1992;	92JP-00325303.
XX	(KAEN)/	KAENNO K.
XX	(SDMO)	SUMITOMO METAL IND. LTD.
XX	(SOYA-)	SOYAKU GIJUTSU KENKYUSHO KK.
XX	WPI: 1995-287962/38.	
XX	N-PSDB; AAF03960.	
XX	Disclosure; Page 39-48; 52pp; Japanese.	
XX	An HCV proteinase active substance - which has activity as an anti-HCV agent and can be used to screen for proteinase inhibitors.	
XX	The present sequence is a partial Hepatitis C Virus (HCV) polyprotein from the non-structural region. Partial proteinase sequences (AAF82692-93) are contained within this sequence. The proteinases can be used as anti-HCV agents. They can also be used to screen cpds. for their ability to inhibit their proteolytic activity. In this way proteinase inhibitors can be identified. (Updated on 16-OCT-2003 to standardise OS field)	
XX	Sequence 3010 AA;	
XX	Query Match	95.8%; Score 1967; DB 2; Length 3010;
XX	Best Local Similarity	93.9%; Pred. No. 3,1e-181;
XX	Matches 369; Conservative	13; Mismatches 11; Indels 0; Gaps 0
QY	1 MAASGGAVFGLTALTTSPYKVLARLIMWLOYLRVEAHLQWTPELNVRGGRDAI	60
DB	8.14 MAASGGAVFGLTALTTSPYKVLARLIMWLOYLRVEAHLQWTPELNVRGGRDAI	873
QY	61 ILTCAVHELLFDITKLLAIFGELMVLQAGITKVFYVPAQGLIRACMLVRKAAGHY	120
DB	8.74 ILTCAVHELLFDITKLLAIFGELMVLQAGITKVFYVPAQGLIRACMLVRKAAGHY	933
QY	121 VQMAFMKLAALTGYTVYDHLTFLQWNAHAGLBDLAVANEPVFSMEVKYITWQADTYAC	180
DB	934 VQMAFMKLAALTGYTVYDHLTFLQWNAHAGLBDLAVANEPVFSMEVKYITWQADTYAC	993
QY	181 GDITIGLPSARSGEILLGPADNFEQGQMRLLAPITAYSQOTRGLGCIITSLGRDKN	240
DB	994 GDITIGLPSARSGEILLGPADNFEQGQMRLLAPITAYSQOTRGLGCIITSLGRDKN	105
QY	241 QYGEVQVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYNNVDQDLVQCA	300
DB	1054 QYGEVQVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYNNVDQDLVQCA	1113
QY	301 PGASGMPCTCGSSDLYLVTHADVIPIRRGDSRGSLSRPYSYLGSSGSPILCS	360
DB	1114 PGASGMPCTCGSSDLYLVTHADVIPIRRGDSRGSLSRPYSYLGSSGSPILCS	117
QY	361 GHAVGIFRAAVCTRGVAKAVDEIPVESMETTNR	393
DB	1174 GHAVGIFRAAVCTRGVAKAVDEIPVESMETTNR	1206

ID	AA68622
ID	AA68622 standard; protein; 3010 AA.
XX	
AC	AA68622;
XX	
DT	16-OCT-2003 (revised)
DT	16-OCT-1995 (first entry)
XX	
DE	HCV protein cleavable with new serine proteinase.
XX	
KM	proteinnase; serine, cleavage; hepatitis C virus; HCV.
OS	Hepatitis C virus; Virus.
FH	Key Location/Qualifiers
FT	Cleavage-site 2419..2420 /note= "Serine protease cleavage site"
XX	
FN	JF06315377-A.
PD	
XX	
PE	15-NOV-1994.
PF	06-MAY-1993; 93JP-00105666.
PR	06-MAY-1993; 93JP-00105666.
PA	(KAENNO K. SUMITOMO METAL IND LTD. SOYAKU GIJUTSU KENKYUSHO KK. WIPI, 1995-033330/05. N-FSDBI; AAO80498.)
DR	
XX	
PT	New HCV-originated proteinase active substance - used for site-specific cleavage by an intermolecular reaction and the purification thereof.
PS	Disclosure; Page 10-19; 23pp; Japanese.
CC	This protein from HCV (hepatitis C virus) (encoded by AAO80498) is cleaved between amino acids 2419 and 2420, by a new serine protease, contg. the sequence of AA68621. The proteinase is purified as a fused product with the dihydrofolate reductase protein by using a methotrexate column. It can be used for the development of an inhibitor for HCV proteinase. (Updated on 16-OCT-2003 to standardise OS field)
XX	
SC	Sequence 3010 AA.
Query Match	95.6%; Score 1962; DB 2; Length 3010; Best Local Similarity 93.6%; Pred.No. 9.4e-18; Matches 368; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
QY	1 MAASCAGAVFGILLLTSPYKYLRLTWLLOYLTRVAHIQWVIPLNVGGRDAT 60
DB	814 MAASCGGAVPGLVLTLSPTKKFLRLRWLOFITRABAHQWVPPLNVGGRDAT 873
QY	61 ILLTCAVHELIPIITKLILAIFGPLMWLAGITKVPYFVAAGLIRACMLVRKAGGHY 120
DB	874 ILLTCAVHELIPIITKLILPIGLFWLQGITRVFPFAAGLIRACMLVRKAGGHY 933
QY	121 VQMAFMKLAALTGYVDHLTPLDKMAHAGLRDAVNEPVIFSDMEVKIITWGADTAA 180
DB	934 VQMAFMKLAALTGYVDHLTPLEDMABAAGLRDAVAEPVFSDEMELKITWGADTAA 993
QY	181 GDIIISGLPVSARRREILGPADNFEGQGMILLAPITAYSQOTSGLGCIITSLTGRDN 240
DB	994 GDIIISGLPVSARRKEILGPSADFEGQGMILLAPITAYSQOTSGLGCIITSLTGRDN 1053
QY	241 QVEGEVOVSTATOSPLATCNVCWTYFHAGSKYTLAHPGPIPTQNTYTNDDLVGNQA 300
DB	1054 QVDEVOVLSTATOSPLATCNVCWTYFHAGSKYTLAHPGPIPTQNTYTNDDLVGNQA 1113
QY	PGARSMPTCCGSSDLYLVTRHADVPVRRRGSRGSLSPRPVSYLKSGSGPLLCP 360

D6		1114	PGARSMTPCTCGSSDYLTVTRHADVPEVRRRGDSRGLSPRPISYKGGSCGPLCPSS	1173
OY		361	GHAVGIFPRAAVCTRGAAKAVDFIPESHMETMR	393
D6		1174	GHWVGIFPRAAVCTRGAAKAVDFIPESHMETMR	1206
RESULT	6			
ID	AAR68864	standard; protein; 3010 AA.		
XX				
AC	AAR68864;			
DI	06-DEC-1995	(first entry)		
XX				
DE	Hepatitis C virus RNA helicase.			
XX				
KM	Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;			
KW	Baculovirus; recombinant production.			
XX				
OS	Hepatitis C virus.			
XX				
FH	Key	Location/Qualifiers		
FT	Region	196..198	/label= N-linked glycosylation site	
FT	Region	209..211	/label= N-linked glycosylation site	
FT	Region	234..236	/label= N-linked glycosylation site	
FT	Region	250..252	/label= N-linked glycosylation site	
FT	Region	305..307	/label= N-linked glycosylation site	
FT	Region	325..327	/label= N-linked glycosylation site	
FT	Region	417..419	/label= N-linked glycosylation site	
FT	Region	423..425	/label= N-linked glycosylation site	
FT	Region	430..432	/label= N-linked glycosylation site	
FT	Region	448..450	/label= N-linked glycosylation site	
FT	Region	532..534	/label= N-linked glycosylation site	
FT	Region	556..558	/label= N-linked glycosylation site	
FT	Region	576..578	/label= N-linked glycosylation site	
FT	Region	623..625	/label= N-linked glycosylation site	
FT	Region	645..647	/label= N-linked glycosylation site	
FT	Region	1213..1215	/label= N-linked glycosylation site	
FT	Region	1255..1257	/label= N-linked glycosylation site	
FT	Region	2041..2043	/label= N-linked glycosylation site	
FT	Region	2077..2079	/label= N-linked glycosylation site	
FT	Region	2240..2242	/label= N-linked glycosylation site	
FT	Region	2788..2790	/label= N-linked glycosylation site	
FT	Region		/label= N-linked glycosylation site	
XX				
PN	JP06319583-A.			
XX				
PD	22-NOV-1994.			
XX				
ZF	18-SEP-1992;	92JF-00249241.		
XX				
RX	18-SEP-1992;	92JF-00249241.		

XX	PA	(SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX	XX	WPI, 1995-040330/06.
DR	DR	N-PSDB; AAQ81559.
XX	XX	of hepatitis C virus helicase gene in baculovirus - useful for large
PT	PT	scale prodn. of RNA helicase.
XX	XX	
PS	PS	Claim 1; Fig 1-4; 3pp; Japanese.
XX	XX	
CC	CC	AAQ81559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The DNA
CC	CC	was used in the construction of an expression vector, which was used to
CC	CC	transform a baculovirus host. The transformed baculovirus could then be
CC	CC	used for the recombinant prodn. of HCV RNA helicase
XX	XX	
XX	XX	Sequence 3010 AA;
QY	QY	Query Match 95.5%; Score 1961; DB 2; Length 3010;
Db	Db	Best Local Similarity 93.6%; Pred. No. 1.2e-180;
		Matches 368; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
QY	QY	1 MAASCGAVFPGTGLTLTSPVYKYLARLIMWLQYLITRYEATQVMIPLINVAGGRDAI 60
Db	Db	814 MAASCGAVFPGTGLTLTSPVYKFLARLIMWLQYFTRBAHLQVWVPLINVAGGRDAI 873
QY	QY	61 ILTCAVHPELIPITKLLALFEGPLAVNLQAGITKVEFYFAQGLIRACMLVRAAGSHY 120
Db	Db	874 ILTCAVHPELIPITKLLALIGPLAVNLQAGITRVFIFYFAQGLIRACMLVRAAGSHY 933
QY	QY	121 VQVAFMKLALTGYVVDHLTPELODMAHAGIRDLAVAVEPIESDMEVKIITWGAOTAAAC 180
Db	Db	934 VQVAFMKLALTGYVVDHLTPELODMAHAGIRDLAVAVEPIESDMEVKIITWGAOTAAAC 993
QY	QY	181 GDIISGIPVSAARGREITLGPADNFEQSGWRILAPITAVSQOTGELGCIITSLTGRDKN 240
Db	Db	994 GDIISGIPVSAARGREITLGPADNFEQSGWRILAPITAVSQOTGELGCIITSLTGRDKN 1053
QY	QY	241 QVEGEVQVSTATQSFATCVNGVCTVFEHAGSKTLAGPKGPIITQMYTNVDDLVNQQA 300
Db	Db	1054 QVDEVEVQVSTATQSFATCVNGVCTVFEHAGSKTLAGPKGPIITQMYTNVDDLVNQQA 1113
QY	QY	301 PPGARSMTPCTCGSSDLYLTVRHADVPEVRRRSGRSGLSPRPVSYLKSSGGGLICPS 360
Db	Db	1114 PPGARSMTPCTCGSSDLYLTVRHADVPEVRRRSGRSGLSPRPVSYLKSSGGGLICPS 1173
QY	QY	361 GHAAGIFFAAVCTRGVAKAVDFIVESMETTMR 393
Db	Db	1174 GHAAGIFFAAVCTRGVAKAVDFIVESMETTMR 1206
RESULT 7		
ABG30601	ID	ABG30601 standard; protein; 2201 AA.
XX	XX	ABG30601;
AC	AC	21-OCT-2002 (first entry)
DT	DT	
XX	XX	Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
XX	XX	Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW	KW	cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.
XX	XX	
OS	OS	Hepatitis C virus.
OS	OS	Synthetic.
XX	XX	
FH	FH	Key Location/Qualifiers
FT	FT	Misc-difference 882
FT	FT	/label= Arg, Lys
FT	FT	Misc-difference 2183
XX	XX	/note= "Wild type Met substituted by Thr"

PN MO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukolj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apk12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7.1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MAASCGAVFTGLALLTSPYKVLRLIWLQYLITRVEALHQLWIPPLNVRGGDAI 60  
 DB 5 MAASCGAVFVGLILLTSPYKVLRLIWLQYFTRAEALHQLWIPPLNVRGGDAV 64  
 QY 61 ILTCAVHPELIPITITLLAIFGPIWVLOAGITKYPYFRAQGLIRACMLVRKAGHY 120  
 DB 65 ILTCAHPELIPITITLLAIFGPIWVLOAGITKYPYFRAQGLIRACMLVRKAGHY 124  
 QY 121 VQNAFKLALVTGYVDHLTPLODWAHAGLRDLAVALVEPVITSDMEVKIITWGADTAAC 180  
 DB 125 VQNALMMLALVTGYVDHLTPLODWAHAGLRDLAVALVEPVITSDMEVKIITWGADTAAC 184  
 QY 181 GDIISGLPVARRGREILGPADNFEQGWRLAPITAYVQQRGLGCTITSLTGRDXN 240  
 DB 185 GDIIIGLPVARRGREILGPADNFEQGWRLAPITAYVQQRGLGCTITSLTGRDXN 244  
 QY 241 QVEGEVQVSTATQSFATCVNGVWTFVHAGSKTLAGEKPIITQWYTNVDDIVGWA 300  
 DB 245 QVEGEVQVSTATQSFATCVNGVWTFVHAGSKTLAGEKPIITQWYTNVDDIVGWA 304  
 QY 301 PPGARSTPTCTGSSDYLVTIRHADVTPVRRRDSGSLSPSPVYIKSSGGPLCP 360  
 DB 305 PPGARSTPTCTGSSDYLVTIRHADVTPVRRRDSGSLSPSPVYIKSSGGPLCP 364  
 QY 361 GHAVGIFRAVCTRGVAKAVDFIPVSMETMR 393  
 DB 365 GHAVGIFRAVCTRGVAKAVDFIPVSMETMR 397

RESULT 8  
 ABG30591  
 ID ABG30591 standard; protein; 2201 AA.  
 XX  
 AC ABG30591;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 751  
 FT /note= "wild type Ser substituted by Gly"  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 PN MO200252015-A2.  
 PD 04-JUL-2002.  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukolj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apk12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7.1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MAASCGAVFTGLALLTSPYKVLRLIWLQYLITRVEALHQLWIPPLNVRGGDAI 60  
 DB 5 MAASCGAVFVGLILLTSPYKVLRLIWLQYFTRAEALHQLWIPPLNVRGGDAV 64  
 QY 61 ILTCAVHPELIPITITLLAIFGPIWVLOAGITKYPYFRAQGLIRACMLVRKAGHY 120

```

Cc      65  ILITCAIHPHLLIFITTKILLALIGPLMVLAQGITKVPYFPAHGLIRACMLVRKAVGSHY 124
Cc      121  VQMAFMKLAALTGTYVDHLTPLODMAHAGLRDLAANAEPVTSDMVEKIIITGADTAAC 180
Cc      125  VQMAFMKLAALTGTYVDHLTPLODMAHAGLRDLAANAEPVTSDMVEKIIITGADTAAC 184
Cc      181  GDIISGLPVARSARREILIGPADNFEQGWRLAPITAYSQQRGLGCIITSLTGRDNX 240
Cc      185  GDIIILGIPVARSARREILIGPADNFEQGWRLAPITAYSQQRGLGCIITSLTGRDNX 244
Cc      241  QVEGEVQVVSATQSFATQVNGVCMVTFHAGSKTLAAGKGPITQMTYNVDQDLVGMQA 300
Cc      245  QVEGEVQVVSATQSFATQVNGVCMVTFHAGSKTLAAGKGPITQMTYNVDQDLVGMQA 304
Cc      301  PPGARSLTPCTCGSSDLYLVTNRADVIPIRRRGRSGSLSPRPVSYLKSGSGPILCPG 360
Cc      305  PPGARSLTPCTCGSSDLYLVTNRADVIPIRRRGRSGSLSPRPVSYLKSGSGPILCPG 364
Cc      361  GHAVGIFRAVCTRGVAKAVDPIPVESMETTMR 393
Cc      365  GHAVGIFRAVCTRGVAKAVDPIPVESMETTMR 397

RESULT 9
ABG30600
ID      ABG30600 standard; protein; 2201 AA.
XX
AC      ABG30600;
XX
DT      21-OCT-2002 (first entry)
XX
DE      Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.
XX
KM      Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
XX      cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.
XX
OS      Hepatitis C virus.
XX
OS      Synthetic.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 882
FT      Misc-difference 1357 Arg, Lys
FT      Misc-difference 1357
FT      Misc-difference 1357 /note= "Wild type Pro substituted by Leu"
XX
XX      MO200252015-A2.
XX
XX      04-JUL-2002.
XX
XX      20-DEC-2001; 2001WO-CA001843.
XX
XX      22-DEC-2000; 2000US-0257857P.
XX
PA      (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI      Kukolj G, Pause A;
XX
XX      WPI; 2002-575382/61.
XX
XX      New self-replicating RNA molecules from Hepatitis C virus (HCV), which
XX      possess enhanced transduction or replication efficiency, useful for
XX      evaluating potential inhibitors of HCV replication.
XX
XX      Claim 3, Page; 140pp; English.
XX
XX      The invention describes a self-replicating hepatitis C virus (HCV)
XX      polynucleotide molecule comprising a 5'-non translated region (NTR),
XX      where guanine at position 1 is substituted for adenine, a HCV polyprotein
XX      region coding for a HCV polypeptide; and a 3'-NTR region. The self-
XX      replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
XX      potential inhibitors of HCV replication. The HCV RNA molecule is also
XX      useful for efficiently establishing cell culture replication. The self-

```

```

Cc      replicating polynucleotide molecule contains a 5'-NTR, where G at
Cc      position 1 is substituted for A, and therefore provides an alternative to
Cc      existing systems comprising a self-replicating HCV RNA molecule that, in
Cc      conjunction with mutations in the HCV non-structural region, such as the
Cc      G(2042)C/R mutations, transduces and/or replicates with greater
Cc      efficiency. This amino acid sequence represents a mutant of the hepatitis
Cc      C virus replicon APGK12 and contains the viral protease NS2/3, protease
Cc      complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
Cc      This sequence does not appear in the specification but has been created
Cc      from the wild type sequence shown in ABG30581 using information given in
Cc      the claims of the invention
Cc
XX      Sequence 2201 AA:
XX
XX      Query Match      95.0%; Score 1951; DB 5; Length 2201;
XX      Best Local Similarity 93.1%; Pred. No. 7, Le-180;
XX      Matches 366; Conservative 14; Mismatches 113; Indels 0; Gaps 0;

Cc      1  MAACGGAVFGLALTLTSPYKVLARLWLOYLITRVEAHQVWIPINVRGGRDAI 60
Cc      5  MAACGGAVFGLALTLTSPYKVLARLWLOYLITRVEAHQVWIPINVRGGRDAI 64
Cc
Cc      61  ILITCAIHPHLLIFITTKILLALIGPLMVLAQGITKVPYFPAHGLIRACMLVRKAVGSHY 120
Cc      65  ILITCAIHPHLLIFITTKILLALIGPLMVLAQGITKVPYFPAHGLIRACMLVRKAVGSHY 124
Cc      121  VQMAFMKLAALTGTYVDHLTPLODMAHAGLRDLAANAEPVTSDMVEKIIITGADTAAC 180
Cc      125  VQMAFMKLAALTGTYVDHLTPLODMAHAGLRDLAANAEPVTSDMVEKIIITGADTAAC 184
Cc      181  GDIISGLPVARSARREILIGPADNFEQGWRLAPITAYSQQRGLGCIITSLTGRDNX 240
Cc      185  GDIIILGIPVARSARREILIGPADNFEQGWRLAPITAYSQQRGLGCIITSLTGRDNX 244
Cc      241  QVEGEVQVVSATQSFATQVNGVCMVTFHAGSKTLAAGKGPITQMTYNVDQDLVGMQA 300
Cc      245  QVEGEVQVVSATQSFATQVNGVCMVTFHAGSKTLAAGKGPITQMTYNVDQDLVGMQA 304
Cc      301  PPGARSLTPCTCGSSDLYLVTNRADVIPIRRRGRSGSLSPRPVSYLKSGSGPILCPG 360
Cc      305  PPGARSLTPCTCGSSDLYLVTNRADVIPIRRRGRSGSLSPRPVSYLKSGSGPILCPG 364
Cc      361  GHAVGIFRAVCTRGVAKAVDPIPVESMETTMR 393
Cc      365  GHAVGIFRAVCTRGVAKAVDPIPVESMETTMR 397

RESULT 10
ABG30581
ID      ABG30581 standard; protein; 2201 AA.
XX
XX      ABG30581;
XX
XX      21-OCT-2002 (first entry)
XX
XX      Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.
XX
XX      Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
XX      cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
XX      Hepatitis C virus.
XX
XX      MO200252015-A2.
XX
XX      04-JUL-2002.
XX
XX      20-DEC-2001; 2001WO-CA001843.
XX
XX      22-DEC-2000; 2000US-0257857P.
XX
PA      (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI      Kukolj G, Pause A;

```

XX WPI; 2002-575382/61.  
 DR N-PSDB; ABK8573.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Disclosure; Page 49-58; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apkx12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
 CC  
 XX  
 SQ Sequence 2201 AA:  
 Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7.1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MAASCGAVFTGIALTLTSPYKYLARLIMWIQYITRVEAHLOWIPPLNVRGGDAI 60  
 DB 5 MAASCGAVFTGIALTLTSPYKYLARLIMWIQYITRVEAHLOWIPPLNVRGGDAV 64  
 QY 61 ILTCAVHPELITPITKLLAIFGRLMVLQAGITKYPYFRAAGLIRACMLVRKAAGHY 120  
 DB 65 ILTCAHPELITPITKLLAIFGRLMVLQAGITKYPYFRAAGLIRACMLVRKAAGHY 124  
 QY 121 VQMAFMKLAALTGYVVDHLTPLODWAHAGRLDAVAVEVIFSDMEVKIITWGADTAA 180  
 DB 125 VQMALMKLAALTGYVVDHLTPLODWAHAGRLDAVAVEVIFSDMEVKIITWGADTAA 184  
 QY 181 GDIISGLPVARSRRREHLGPADNFBQGRMLAPITAYSQQRGLGCIITSLTGRDXN 240  
 DB 185 GDIIIGLPVARSRRREHLGPADNFBQGRMLAPITAYSQQRGLGCIITSLTGRDXN 244  
 QY 241 QVEGEVQVSTATQSPFLATCUNGCWTVFAGSKTLAGKGPITQMTYNVDODLVGWA 300  
 DB 245 QVEGEVQVSTATQSPFLATCUNGCWTVFAGSKTLAGKGPITQMTYNVDODLVGWA 304  
 QY 301 PPGARSMTPTCTGSSDLYLTRHADVIYVRRRCDGSLSPRPVYKSSGGLTCS 360  
 DB 305 PPGARSMTPTCTGSSDLYLTRHADVIYVRRRCDGSLSPRPVYKSSGGLTCS 364  
 QY 361 GHAVGIFRAVCTRGVAKAVDFVPSMETTMR 393  
 DB 365 GHAVGIFRAVCTRGVAKAVDFVPSMETTMR 397  
 RESULT 11  
 ABG30593  
 ID ABG30593 standard; protein; 2201 AA.  
 XX  
 AC ABG30593;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #4.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /Label= Arg, Lys  
 FT Misc-difference 892  
 FT /note= "Wild type Leu substituted by Phe"  
 XX  
 PN W0200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apkx12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 XX  
 SQ Sequence 2201 AA:  
 Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7.1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MAASCGAVFTGIALTLTSPYKYLARLIMWIQYITRVEAHLOWIPPLNVRGGDAI 60  
 DB 5 MAASCGAVFTGIALTLTSPYKYLARLIMWIQYITRVEAHLOWIPPLNVRGGDAV 64  
 QY 61 ILTCAVHPELITPITKLLAIFGRLMVLQAGITKYPYFRAAGLIRACMLVRKAAGHY 120  
 DB 65 ILTCAHPELITPITKLLAIFGRLMVLQAGITKYPYFRAAGLIRACMLVRKAAGHY 124  
 QY 121 VQMAFMKLAALTGYVVDHLTPLODWAHAGRLDAVAVEVIFSDMEVKIITWGADTAA 180  
 DB 125 VQMALMKLAALTGYVVDHLTPLODWAHAGRLDAVAVEVIFSDMEVKIITWGADTAA 184  
 QY 181 GDIISGLPVARSRRREHLGPADNFBQGRMLAPITAYSQQRGLGCIITSLTGRDXN 240  
 DB 185 GDIIIGLPVARSRRREHLGPADNFBQGRMLAPITAYSQQRGLGCIITSLTGRDXN 244  
 QY 241 QVEGEVQVSTATQSPFLATCUNGCWTVFAGSKTLAGKGPITQMTYNVDODLVGWA 300

Db 245 QVEGEVQVSTATQSFATCVNGVCMVTHGAGSKTLGPKGPIITQMYTNVDQDLVGWQA 304  
QY 301 PPGARSMPTCTGSSDYLVTNRHADVIPIVRRGDSRGSLSPPRYSLKSSGGPILCP8 360  
Db 305 PPGARSLTPCTGSSDYLVTNRHADVIPIVRRGDSRGSLSPPRYSLKSSGGPILCP8 364  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397

RESULT 12  
ABG30582  
ID ABG30582 standard; protein; 2201 AA.  
AC ABG30582;  
XX 21-OCT-2002 (first entry)  
XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.  
XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
XX cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
XX Hepatitis C virus.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 882 /note= "Wild type Lys substituted by Lys or Arg"  
FT Misc-difference 1233 /note= "Wild type Gly substituted by Cys"  
FT  
XX  
XX WO200252015-A2.  
XX 04-JUL-2002.  
XX 20-DEC-2001; 2001WO-CA001843.  
XX 22-DEC-2000; 2000US-0257857P.  
XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX Kukulj G, Pause A;  
XX WPI; 2002-575382/61.  
XX N-PSDB; ABK88574.  
XX  
XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
XX possess enhanced transduction or replication efficiency, useful for  
XX evaluating potential inhibitors of HCV replication.  
XX  
XX Disclosure; Page 59-69; 140pp; English.  
XX  
XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polypeptide  
XX region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in  
XX conjunction with mutations in the HCV non-structural region, such as the  
XX G(2042)/C/R mutations, transduces and/or replicates with greater  
XX efficiency. This amino acid sequence is encoded by the hepatitis C virus  
XX NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
XX sequence has been created from replicon APGX12 shown in ABG30581  
XX  
XX Sequence 2201 AA;  
XX  
XX Query Match 95.0%; Score 1951; DB 5; Length 2201;

Best Local Similarity 93.1%; Pred. No. 7,1e-180;  
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
QY 1 MAASCGAVFGLATLTSPYKVLARLWQLVLRVRAHLOVTPPLNVGGPDAI 60  
Db 5 MAASCGAVFVGLILTTSPHYKPLFARLWQLVLRVRAHLOVTPPLNVGGPDAV 64  
QY 61 ILTCAVHELFIDITKLALFGBLWLOAGITVVPVFAQGLIRACMLVRKAAGHY 120  
Db 65 ILTCAVHELFITTKLILALGFLWLOAGITVVPVFAHAGLIRACMLVRKAAGHY 124  
QY 121 VQMAFKALALTGTIVYVHLPLODMANAGRLDAVAVPEVTFESMEKIIITWGDIAAC 180  
Db 125 VQMALMKALALTGTIVYVHLPLODMANAGRLDAVAVPEVTFESMEKIIITWGDIAAC 184  
QY 181 GDIIISGLPVASARRREILGPADNFEQGGWRLAPITAYSOOTRGLGCIITSLTGRDKN 240  
Db 185 GDIIISGLPVASARRREILGPADNFEQGGWRLAPITAYSOOTRGLGCIITSLTGRDKN 244  
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTHGAGSKTLGPKGPIITQMYTNVDQDLVGWQA 300  
Db 245 QVEGEVQVSTATQSFATCVNGVCMVTHGAGSKTLGPKGPIITQMYTNVDQDLVGWQA 304  
QY 301 PPGARSMPTCTGSSDYLVTNRHADVIPIVRRGDSRGSLSPPRYSLKSSGGPILCP8 360  
Db 305 PPGARSLTPCTGSSDYLVTNRHADVIPIVRRGDSRGSLSPPRYSLKSSGGPILCP8 364  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397

RESULT 13  
ABG30580  
ID ABG30580 standard; protein; 2201 AA.  
AC ABG30580;  
XX 21-OCT-2002 (first entry)  
XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.  
XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
XX cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
XX Hepatitis C virus.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 882 /note= "Encoded by ARG"  
FT  
XX  
XX WO200252015-A2.  
XX 04-JUL-2002.  
XX 20-DEC-2001; 2001WO-CA001843.  
XX 22-DEC-2000; 2000US-0257857P.  
XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX Kukulj G, Pause A;  
XX WPI; 2002-575382/61.  
XX  
XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
XX possess enhanced transduction or replication efficiency, useful for  
XX evaluating potential inhibitors of HCV replication.  
XX  
XX Disclosure; Page 69-74; 140pp; English.  
XX  
XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX



CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apgk12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

XX Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;

Best Local Similarity 93.1%; Pred. No. 7.1e-180; Mismatches 13; Indels 0; Gaps 0;

Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIMLQYITRVEAHQVWIPPLNVRGGRDAI 60  
 DB 5 MAASCGAVFVGLALTLSPYKVLARLIMLQYITRVEAHQVWIPPLNVRGGRDAV 64  
 QY 61 ILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
 DB 65 ILITCAHPELIFITKLLAIFGRLMVLQAGITKVPYFRAQGLIRACMLVRKAAGHY 124  
 QY 121 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPFVSDMEVKTITWADTAAC 180  
 DB 125 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPFVSDMEVKTITWADTAAC 184  
 QY 181 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQOIRGLIGCIITSLTGRDN 240  
 DB 185 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQOIRGLIGCIITSLTGRDN 244  
 QY 241 QVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPKGPITQMTYTNVDDLVGMOA 300  
 DB 245 QVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPKGPITQMTYTNVDDLVGMOA 304  
 QY 301 PPGARSMTPCTCGSSDLYLTVRHADVIPVRRGDSRGLSPRVSYLKSSGGPILCP 360  
 DB 305 PPGARSLTPCTCGSSDLYLTVRHADVIPVRRGDSRGLSPRVSYLKSSGGPILCP 364  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
 DB 365 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 397

RESULT 14

ABG30587  
 ID ABG30587 standard; protein; 2201 AA.

XX AC ABG30587;  
 XX DT 21-OCT-2002 (first entry)  
 XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #7.  
 XX KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 XX KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 XX OS Hepatitis C virus.  
 XX PN WO200252015-A2.  
 XX PD 04-JUL-2002.  
 XX PF 20-DEC-2001; 2001MO-CA001843.  
 XX PR 22-DEC-2000; 2000US-0257657P.  
 XX PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.

XX Kukulj G, Pause A;  
 PI WPI; 2002-575382/61.  
 DR N-Psdb; ABK88587.  
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.

PS Disclosure; Page 120-129; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apgk12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

XX Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;

Best Local Similarity 93.1%; Pred. No. 7.1e-180; Mismatches 13; Indels 0; Gaps 0;

Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIMLQYITRVEAHQVWIPPLNVRGGRDAI 60  
 DB 5 MAASCGAVFVGLALTLSPYKVLARLIMLQYITRVEAHQVWIPPLNVRGGRDAV 64  
 QY 61 ILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
 DB 65 ILITCAHPELIFITKLLAIFGRLMVLQAGITKVPYFRAQGLIRACMLVRKAAGHY 124  
 QY 121 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPFVSDMEVKTITWADTAAC 180  
 DB 125 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPFVSDMEVKTITWADTAAC 184  
 QY 181 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQOIRGLIGCIITSLTGRDN 240  
 DB 185 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQOIRGLIGCIITSLTGRDN 244  
 QY 241 QVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPKGPITQMTYTNVDDLVGMOA 300  
 DB 245 QVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPKGPITQMTYTNVDDLVGMOA 304  
 QY 301 PPGARSMTPCTCGSSDLYLTVRHADVIPVRRGDSRGLSPRVSYLKSSGGPILCP 360  
 DB 305 PPGARSLTPCTCGSSDLYLTVRHADVIPVRRGDSRGLSPRVSYLKSSGGPILCP 364  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
 DB 365 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 397

RESULT 15

ABG30599  
 ID ABG30599 standard; protein; 2201 AA.

XX AC ABG30599;  
 XX DT 21-OCT-2002 (first entry)  
 XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #8.

KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutin.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 892 /label= Arg, Lys  
 FT Misc-difference 1346 /note= "wild type leu substituted by pro"  
 FT  
 XX  
 XX WO200252015-A2.  
 PN  
 PD 04-JUL-2002.  
 XX  
 XX 20-DEC-2001; 2001WO-CA001843.  
 XX  
 XX 22-DEC-2000; 2000US-0257857P.  
 XX  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA  
 XX Kukulj G, Pause A;  
 PI  
 XX  
 XX WPI; 2002-575382/61.  
 DR  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 PT  
 XX  
 XX  
 PS Claim 3; Page; 140pp; English.  
 CC  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon APOX12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 XX  
 XX  
 SQ Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7.1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGGAVFGLALLTSPYKVLARLIMWLQYLITRVAHLQVWIPPLNVRGGDAI 60  
 DB 5 MAASCGGAVFGLALLTSPYKVLARLIMWLQYLITRVAHLQVWIPPLNVRGGDAV 64  
 QY 61 ILTCAHPELIPITITLLAIFGPMVVOAGITKVPYFRAOGLRACMLVRKAGGHY 120  
 DB 65 ILTCAHPELIPITITLLAIFGPMVVOAGITKVPYFRAOGLRACMLVRKAGGHY 124  
 QY 121 VQNAFMLAALTGTYYVDHTPLQDMNAAGLRDLAAVEVIFSDMEVKIITWGAADTAAC 180  
 DB 125 VQNAFMLAALTGTYYVDHTPLQDMNAAGLRDLAAVEVIFSDMEVKIITWGAADTAAC 184  
 QY 181 GDIISGLPVGARRGRELILGPADNFEQGWRLAPITAYSQOTRGLIGCITISLGRDKY 240  
 DB 185 GDIISGLPVGARRGRELILGPADNFEQGWRLAPITAYSQOTRGLIGCITISLGRDKY 244

QY 241 OVEGEVQVASTATOSFLATCVNGVCTVPHGAGSKTLGPKPITOMYTNVDQDLVGKQA 300  
 DB 245 OVEGEVQVASTATOSFLATCVNGVCTVPHGAGSKTLGPKPITOMYTNVDQDLVGKQA 304  
 QY 301 PEGARSMTPTCGSSDLVLTTRHADVIPRRRGDSRGSLSPPVSYLKSSGGPPLCP 360  
 DB 305 PEGARSMTPTCGSSDLVLTTRHADVIPRRRGDSRGSLSPPVSYLKSSGGPPLCP 364  
 QY 361 GHAAGIFRAAVCTRGVAKAVDPFIPVSMETMR 393  
 DB 365 GHAAGIFRAAVCTRGVAKAVDPFIPVSMETMR 397

Search completed: May 6, 2004, 09:30:45  
 Job time : 54.493 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 12.7992 Seconds  
(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-11

Sequence: 1 MAASCGAVFIGLALITLSP.....RGVAKAVDPFVPSVETMTMR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: PIR\_78:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1967	95.8	3010	1 GNMVCU	genome polypeptide
2	1950	95.0	3010	1 A45573	genome polypeptide
3	1935	94.3	3010	1 GNMVW	genome polypeptide
4	1910	93.0	3010	1 S18030	genome polypeptide
5	1888	92.0	3010	1 GNMVTC	genome polypeptide
6	1764	85.9	3011	1 GNMVW3	genome polypeptide
7	1758	85.6	3011	1 S40770	genome polypeptide
8	1752	85.3	3011	1 GNMVCH	genome polypeptide
9	1493	72.7	3014	1 JG5620	genome polypeptide
10	1403	68.3	3033	1 J01303	genome polypeptide
11	1401	68.2	3033	1 GNMVU8	genome polypeptide
12	408.5	19.9	3005	2 T08841	polypeptide - dour
13	342.5	16.7	2970	2 T08839	polypeptide - marm
14	112	5.5	692	2 H71426	hypothetical prote
15	102.5	5.0	660	2 VHWMH2	structural protein
16	101.5	4.9	3434	1 GNMVW	genome polypeptide
17	101	4.9	564	2 S36637	signal recognition
18	101	4.9	600	2 A46642	DNA-directed DNA p
19	100.5	4.9	353	2 G87392	conserved hypotet
20	99	4.8	399	2 AH3038	conserved hypotet
21	99	4.8	399	2 G98247	hypothetical 50.8k
22	99	4.8	446	2 AF1509	conserved hypotet
23	99	4.8	451	2 H82044	C4-dicarboxylate t
24	97.5	4.7	1085	2 T03531	cohn protein homol
25	95.5	4.7	470	2 JG4038	tetracycline 6-hyd
26	94.5	4.6	904	2 A84212	hypothetical prote
27	94.5	4.6	2796	2 JG4743	fatty-acid synthas
28	94	4.6	434	2 G82728	conserved hypotet
29	93.5	4.6	418	2 H90679	probable transport

30	93.5	4.6	418	2 D85530	probable transport
31	93.5	4.6	477	2 H75026	oligopeptide abc t
32	93.5	4.6	1380	2 T18309	receptor-adenylate
33	93.5	4.6	3069	2 H70656	fatty-acid synthas
34	93	4.5	3414	1 GNMVNE	genome polypeptide
35	93	4.5	7463	2 T36248	CDA peptide synthe
36	92.5	4.5	665	2 D83252	nucleotide sugar e
37	92.5	4.5	706	2 S33761	transferrin precu
38	92.5	4.5	716	2 G83612	hypothetical prote
39	92	4.5	659	1 B44212	structural protein
40	91.5	4.5	401	1 A36961	pllin biogenesis p
41	91.5	4.5	428	2 AF0241	probable coenzyme
42	91.5	4.5	446	2 A31150	conserved hypotet
43	91.5	4.5	3412	1 GNMVTE	genome polypeptide
44	90.5	4.4	418	2 A64765	probable transport
45	90.5	4.4	868	2 H81775	aconitate hydratase

## ALIGNMENTS

### RESULT 1

GNMVCU genome polypeptide - hepatitis C virus (strain J)

N:contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Jun-1992 #sequence\_revision 30--Jun-1992 #text\_change 19-Jun-2001

C/Accession: A39253; PS0086

R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients v

A:Reference number: A39253; MUID:91088550; PMID:2175903

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 <KAT>

A:Cross-references: GB:D90208; NID:G221610; PID:BA1423.1; PID:G221611

R:Kato, N.; Ohkoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari

A:Reference number: PS0085

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 <KA2>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polypeptide have not been determined.

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine

F:2-115/Product: capsid protein C #status predicted <CPC>

F:16-191/Product: envelope protein M #status predicted <EM>

F:192-389/Product: major envelope protein E #status predicted <MEB>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitisin #status predicted <NS3>

F:2230-1337/Region: nucleotide-binding motif A (P-loop)

F:312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NA5>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2;

Query Match 95.8%; Score 1967; DB 1; Length 3010;

Best Local Similarity 93.9%; Pred. No. 1.8e-153;

Matches 369; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

DB 1 MAASCGAVFIGLALITLSPYKVLARLIMWLQYLIRVEAHLOWTIPINVRGGRDAI 60

DB 814 MAASCGAVFVGLVLLTISPYKVFLLARLIMWLQYLIRVEAHLOWTIPINVRGGRDAI 873

DB 61 ILLTCAVPELIFDTIKLLAIFGPLYLQAGITKVPYFVRAQGLIRACMLVRKAGGHY 120

DB 874 ILLTCAVPELIFDTIKLLAIFGPLYLQAGITKVPYFVRAQGLIRACMLVRKAGGHY 933

QY 121 VQMAFKALALTGTYVDHLTPLODMAHAGRLDAVAEPIVFSDEVEKITMGADTAAC 180  
 DB 934 VQMAFKALALTGTYVDHLTPLODMAHAGRLDAVAEPIVFSDEVEKITMGADTAAC 993  
 QY 181 GDIISGLPVASARRREILGADNFEQGWRLAPITAVSQQRGLGCIITSLGRDKN 240  
 DB 994 GDIISGLPVASARRREILGADNFEQGWRLAPITAVSQQRGLGCIITSLGRDKN 1053  
 QY 241 QVEGEVQVAVSTATOSFLATCNGVCMWTFHAGSKTLAGEKPIITQMTYNVDQDLVGMQA 300  
 DB 1054 QVEGEVQVAVSTATOSFLATCNGVCMWTFHAGSKTLAGEKPIITQMTYNVDQDLVGMQA 1113  
 QY 301 PPGARSTPCTCGSSDLYLTRHADVIPIVRRGDSRSLSPRVSYLKSSGGPILCP 360  
 DB 1114 PPGARSTPCTCGSSDLYLTRHADVIPIVRRGDSRSLSPRVSYLKSSGGPILCP 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
 DB 1174 GHVIGIFRAVCTRGVAKAVDFIPVESMETMR 1206

## RESULT 2

A45573  
 genome polyprotein - hepatitis C virus (strain J7)  
 N:Contains: capsid protein C; envelope protein M; hepatitis virin (EC 3.4.21.98) (nonstructu  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: A45573  
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,  
 Virus Res. 23, 39-53, 1992  
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: B  
 A:Reference number: A45573; PMID:9225714; PMID:1318627  
 A:Accession: A45573  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3010 <TAN>  
 A:Cross-references: GB:D11168; GB:D01171; NID:G221612; PIDN:BA01943.1; PID:G221613  
 A:Experimental source: HCV-JT  
 A:Note: Sequence extracted from NCBI backbone (NCIN:106206, NCBI:106207)  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <NEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis virin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1316-1862/Product: nonstructural protein NS4 #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 95.0%; Score 1950; DB 1; Length 3010;

Best Local Similarity 93.6%; Pred. No. 4.7e-152;

Matches 369; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 MAASCGAVFVIGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFVIGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 874 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGGHY 933  
 QY 121 VQMAFKALALTGTYVDHLTPLODMAHAGRLDAVAEPIVFSDEVEKITMGADTAAC 180  
 DB 934 VQMAFKALALTGTYVDHLTPLODMAHAGRLDAVAEPIVFSDEVEKITMGADTAAC 993  
 QY 181 GDIISGLPVASARRREILGADNFEQGWRLAPITAVSQQRGLGCIITSLGRDKN 240

DB 994 GDIISGLPVASARRREILGADNFEQGWRLAPITAVSQQRGLGCIITSLGRDKN 1053  
 QY 241 QVEGEVQVAVSTATOSFLATCNGVCMWTFHAGSKTLAGEKPIITQMTYNVDQDLVGMQA 300  
 DB 1054 QVEGEVQVAVSTATOSFLATCNGVCMWTFHAGSKTLAGEKPIITQMTYNVDQDLVGMQA 1113  
 QY 301 PPGARSTPCTCGSSDLYLTRHADVIPIVRRGDSRSLSPRVSYLKSSGGPILCP 360  
 DB 1114 PPGARSTPCTCGSSDLYLTRHADVIPIVRRGDSRSLSPRVSYLKSSGGPILCP 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
 DB 1174 GHVIGIFRAVCTRGVAKAVDFIPVESMETMR 1206

## RESULT 3

GNMVTM  
 genome polyprotein - hepatitis C virus (strain Taiwan)  
 N:Contains: capsid protein C; envelope protein M; hepatitis virin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A40244  
 R:Chen, P.T.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
 Virology 188, 102-113, 1992  
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
 A:Reference number: A40244; PMID:92230206; PMID:1314449  
 A:Accession: A40244  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <CHE>  
 A:Cross-references: GB:M64754  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:116-191/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <NEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis virin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,223,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 94.3%; Score 1935; DB 1; Length 3010;

Best Local Similarity 91.9%; Pred. No. 8.1e-151;

Matches 361; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFVIGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFVIGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 874 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGGHY 933  
 QY 121 VQMAFKALALTGTYVDHLTPLODMAHAGRLDAVAEPIVFSDEVEKITMGADTAAC 180  
 DB 934 VQMAFKALALTGTYVDHLTPLODMAHAGRLDAVAEPIVFSDEVEKITMGADTAAC 993  
 QY 181 GDIISGLPVASARRREILGADNFEQGWRLAPITAVSQQRGLGCIITSLGRDKN 240  
 DB 994 GDIISGLPVASARRREILGADNFEQGWRLAPITAVSQQRGLGCIITSLGRDKN 1053  
 QY 241 QVEGEVQVAVSTATOSFLATCNGVCMWTFHAGSKTLAGEKPIITQMTYNVDQDLVGMQA 300  
 DB 1054 QVEGEVQVAVSTATOSFLATCNGVCMWTFHAGSKTLAGEKPIITQMTYNVDQDLVGMQA 1113

QY 301 PGARSMPTCTCGSSDLYLVTRHADVTVPARRGDSRGSLLSPRPVSYLKSSGGGFLCPG 360  
 DB 1114 PGARSLTPTCTCGSSDLYLVTRHADVTVPARRGDSRGSLLSPRPVSYLKSSGGGFLCPG 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 1206

## RESULT 4

S18030  
 genome polypeptide - hepatitis C virus (isolate JX1)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Variety: isolate JX1  
 C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
 C/Accession: S18030; S3570; A48332; S18029  
 R/Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 Submitted to the EMBL Data Library, September 1991  
 A/Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
 A/Reference number: S18028  
 A/Accession: S18030  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <HON>  
 A/Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479  
 A/Experimental source: isolate JX1 from an individual  
 R/Honda, M.; Kaneko, S.; Ueno, M.; Kobayashi, K.; Murakami, S.  
 Arch. Virol. 128, 163-169, 1993  
 A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A/Reference number: A48332; MUID:93119270; PMID:8380322  
 A/Accession: S33570  
 A/Molecule type: genomic RNA  
 A/Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>  
 A/Cross-references: EMBL:X61591  
 A/Note: this sequence is inconsistent with the nucleotide translation  
 A/Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser  
 A/Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)  
 C/Superfamily: hepatitis C virus genome polypeptide  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
 F:116-119/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 93.0%; Score 1910; DB 1; Length 3010;  
 Best Local Similarity 91.8%; Pred. No. 9.3e-149;  
 Matches 361; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 1 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHLOVWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHLOVWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVHPELFDITKLLAIFGLMTLOAGITKVPYFAOGLIRACMLVRKAAGHY 120  
 DB 874 ILITCAVHPELFDITKLLAIFGLMTLOAGITKVPYFAOGLIRACMLVRKAAGHY 933  
 QY 121 VQMAFMKALALGTGVYVDHLTPLODMAHAGRLDAVAEVPYFSMEVKITTWGADTAAC 180  
 DB 934 VQMAFMKALALGTGVYVDHLTPLODMAHAGRLDAVAEVPYFSMEVKITTWGADTAAC 993  
 QY 181 GDIIISGLPVASARREIILGPADNFEQGMFLAPITVYSQOTRGLGCIITISLGRDXN 240

DB 994 GDIIISGLPVASARREIILGPADNFEQGMFLAPITVYSQOTRGLGCIITISLGRDXN 1053  
 QY 241 QVEGEVQVSTATQGFPLATCVNGVCMVPHGAGSKTLGAPGPITOMTNTVDDIVGQA 300  
 DB 1054 QVEGEVQVSTATQGFPLATCVNGVCMVPHGAGSKTLGAPGPITOMTNTVDDIVGQA 1113  
 QY 301 PGARSMPTCTCGSSDLYLVTRHADVTVPARRGDSRGSLLSPRPVSYLKSSGGGFLCPG 360  
 DB 1114 PGARSLTPTCTCGSSDLYLVTRHADVTVPARRGDSRGSLLSPRPVSYLKSSGGGFLCPG 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 1206

## RESULT 5

GNVATC  
 genome polypeptide - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
 C/Accession: A38465  
 R/Takamizawa, A.; Mori, C.; Fukey, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991  
 A/Title: Structure and organization of the hepatitis C virus genome isolated from human  
 A/Reference number: A38465; MUID:91140698; PMID:1847440  
 A/Accession: A38465  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <TKX>  
 A/Cross-references: EMBL:M58335; NID:G329770; PIDN:AA72945.1; PID:G329771  
 C/Superfamily: hepatitis C virus genome polypeptide  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:116-119/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 92.0%; Score 1888; DB 1; Length 3010;  
 Best Local Similarity 90.8%; Pred. No. 6.1e-147;  
 Matches 357; Conservative 13; Mismatches 23; Indels 0; Gaps 0;

QY 1 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHLOVWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHLOVWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVHPELFDITKLLAIFGLMTLOAGITKVPYFAOGLIRACMLVRKAAGHY 120  
 DB 874 ILITCAVHPELFDITKLLAIFGLMTLOAGITKVPYFAOGLIRACMLVRKAAGHY 933  
 QY 121 VQMAFMKALALGTGVYVDHLTPLODMAHAGRLDAVAEVPYFSMEVKITTWGADTAAC 180  
 DB 934 VQMAFMKALALGTGVYVDHLTPLODMAHAGRLDAVAEVPYFSMEVKITTWGADTAAC 993  
 QY 181 GDIIISGLPVASARREIILGPADNFEQGMFLAPITVYSQOTRGLGCIITISLGRDXN 240  
 DB 994 GDIIISGLPVASARREIILGPADNFEQGMFLAPITVYSQOTRGLGCIITISLGRDXN 1053  
 QY 241 QVEGEVQVSTATQGFPLATCVNGVCMVPHGAGSKTLGAPGPITOMTNTVDDIVGQA 300  
 DB 1054 QVEGEVQVSTATQGFPLATCVNGVCMVPHGAGSKTLGAPGPITOMTNTVDDIVGQA 1113  
 QY 301 PGARSMPTCTCGSSDLYLVTRHADVTVPARRGDSRGSLLSPRPVSYLKSSGGGFLCPG 360

Db 1114 PGASRLPCTCGSSDLYLTVTRADVI PVRRRGRSGSLSPRPVSYLKGSSGGLCPA 1173  
QY 361 GHAAGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
Db 1174 GHAAGIFRAAVCTRGVAKAVDFIPVESMETMR 1206

## RESULT 6

GNVWC3  
genome polypeptide - hepatitis C virus (strain HCV-1)  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #ext\_change 19-Jan-2001  
C:Accession: A39166; PQ0403; PQ0404  
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co  
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A:Title: Genetic organization and diversity of the hepatitis C virus.  
A:Reference number: A39166; PMID:91172826; PMID:148704  
A:Accession: A39166  
A:Molecule type: mRNA  
A:Residues: 1-3011 <CHO>  
A:Cross-references: GB:M62321; NID:9329873; PIDN:AAA5676.1; PID:G329874  
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L  
U. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
A:Reference number: PQ0393; PMID:9226871; PMID:1316939  
A:Accession: PQ0403  
A:Molecule type: genomic RNA  
A:Residues: 1577-1633 <CHA>  
A:Cross-references: DDB:DJ0128  
A:Experimental source: Isolates E-b16  
A:Accession: PQ0404  
A:Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1577-1633 <CH2>  
A:Experimental source: Isolates E-b17  
A:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolyase; nonstructu  
F:115/115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:132-389/Product: major envelope protein E #status predicted <ME>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1635/Product: hepatitis virus #status predicted <NS3>  
F:1230-1237/Product: nucleotide-binding motif A (P-loop)  
F:1312-1319/Region: DEXH motif  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:136,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 85.9%; Score 1764; DB 1; Length 3011;  
Best Local Similarity 81.9%; Pred. No. 1e-136;  
Matches 32; Conservative 34; Mismatches 37; Indels 0; Gaps 0;  
QY 1 MAASCGAVFGLALTLSPYKVLARLIWLTQYLITRVAHLQVPIPLNVRGGDAI 60  
Db 814 VAASCGGVVAVGMAALTLSPYKRLISWCLMWLFYTRVBAQLHWVWPPPLNVRGGDAV 873  
QY 61 ILTCAVHEPILPITKLLAIFGLPLVLAQGITKVPYFVRAQGLIRACMLVKAAGHY 120  
Db 874 ILTCAVHEPILPITKLLAIFGLPLVLAQGITKVPYFVRAQGLIRACMLVKAAGHY 933  
QY 121 VQAFMKLALTLGVYVDHLPLODMAHAGLRDLAVAVEPIFSDEVKITITWGAADTAAC 180  
Db 934 VQAFMKLALTLGVYVDHLPLODMAHAGLRDLAVAVEPIFSDEVKITITWGAADTAAC 993  
QY 181 GDIISGLPVSAARRREITLGPADNFEQGRRLAPITAYSQOTRGILGCIITSLTGSDKN 240  
Db 994 GDIISGLPVSAARRREITLGPADNFEQGRRLAPITAYSQOTRGILGCIITSLTGSDKN 1053

QY 241 QVEGEVQVSTATQSFPLATVNGVCWTFVHGAGSKTLACXKPIITQNTTWNDDIYGMQA 300  
Db 1054 QVEGEVQVSTATQSFPLATVNGVCWTFVHGAGSKTLACXKPIITQNTTWNDDIYGMQA 1113  
QY 301 PGASRLPCTCGSSDLYLTVTRADVI PVRRRGRSGSLSPRPVSYLKGSSGGLCPA 360  
Db 1114 PGASRLPCTCGSSDLYLTVTRADVI PVRRRGRSGSLSPRPVSYLKGSSGGLCPA 1173  
QY 361 GHAAGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
Db 1174 GHAAGIFRAAVCTRGVAKAVDFIPVESMETMR 1206

## RESULT 7

540770  
genome polypeptide - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence revision 19-May-2000 #ext\_change 19-Jan-2001  
C:Accession: S40770; PC1285  
R:Okamoto, H.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S40770  
A:Accession: S40770  
A:Molecule type: genomic RNA  
A:Residues: 1-3011 <OKA>  
A:Cross-references: EMBL:DJ0749; NID:9221586; PIDN:BA01582.1; PID:9221587  
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,  
Jpn. J. Exp. Med. 60, 167-177, 1990  
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.  
A:Reference number: PC1284; PMID:9101116; PMID:2170712  
A:Accession: PC1285  
A:Molecule type: genomic RNA  
A:Residues: 1-513 <OK2>  
A:Cross-references: GB:D00831; NID:9221511; PIDN:BA00705.1; PID:9221512  
A:Experimental source: isolate HC-01  
A:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polypeptide; serin  
F:2115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <ME>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1635/Product: hepatitis virus #status predicted <NS3>  
F:1230-1237/Product: nucleotide-binding motif A (P-loop)  
F:1312-1319/Region: DEXH motif  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 85.6%; Score 1758; DB 1; Length 3011;  
Best Local Similarity 81.4%; Pred. No. 3.2e-136;  
Matches 32; Conservative 34; Mismatches 39; Indels 0; Gaps 0;  
QY 1 MAASCGAVFGLALTLSPYKVLARLIWLTQYLITRVAHLQVPIPLNVRGGDAI 60  
Db 814 VAASCGGVVAVGMAALTLSPYKRLISWCLMWLFYTRVBAQLHWVWPPPLNVRGGDAV 873  
QY 61 ILTCAVHEPILPITKLLAIFGLPLVLAQGITKVPYFVRAQGLIRACMLVKAAGHY 120  
Db 874 ILTCAVHEPILPITKLLAIFGLPLVLAQGITKVPYFVRAQGLIRACMLVKAAGHY 933  
QY 121 VQAFMKLALTLGVYVDHLPLODMAHAGLRDLAVAVEPIFSDEVKITITWGAADTAAC 180  
Db 934 VQAFMKLALTLGVYVDHLPLODMAHAGLRDLAVAVEPIFSDEVKITITWGAADTAAC 993  
QY 181 GDIISGLPVSAARRREITLGPADNFEQGRRLAPITAYSQOTRGILGCIITSLTGSDKN 240  
Db 994 GDIISGLPVSAARRREITLGPADNFEQGRRLAPITAYSQOTRGILGCIITSLTGSDKN 1053  
QY 241 QVEGEVQVSTATQSFPLATVNGVCWTFVHGAGSKTLACXKPIITQNTTWNDDIYGMQA 300



```
Db 1054 QVEGEVQIVSTATQTFPLATCINGVCMTVYHGAGTFTIASPKGPVLIQTYTNVDOLVGNPA 1113
QY 301 PGARSMTPTCTGSSDLYLVTRHADVI PVRRGDSRGSLSLSPRVSYLKSSGGPILCPG 360
Db 1114 PGARSMTPTCTGSSDLYLVTRHADVI PVRRGDSRGSLSLSPRVSYLKSSGGPILCPG 1173
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 393
Db 1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206
```

## RESULT 8

```
GENWYCH
genome polypeptide - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
Protein NS4; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Accession: A36814; A41546
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
R:Inchuspe, G.; Zebdeed, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Submitted to Genbank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3014 <N>
A:Cross-references: GB:M67463; NID:G329737; PIND:AAA45534.1; PID:G329738
R:Inchuspe, G.; Zebdeed, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructu
F:1.115/Product: capsid protein C #status predicted <CPC>
F:1.115/Product: envelope protein M #status predicted <EMP>
F:1.191/Product: major envelope protein E #status predicted <MEB>
F:1.389/Product: nonstructural protein NS1 #status predicted <NS1>
F:1.390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F:1.707-1006/Product: nonstructural protein NS3 #status predicted <NS3>
F:1.1007-1616/Product: hepatitisin #status predicted <NS4>
F:1.130-1237/Product: nucleotide-binding motif A (P-loop)
F:1.1312-1317/Region: nucleotide-binding motif B
F:1.1316-1319/Region: DEXH motif
F:1.1616-1863/Product: nonstructural protein NS4 #status predicted <NS4>
F:1.1663-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2.014-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:1.196,209,334,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23
```

Query Match 85.3%; Score 1752; DB 1; Length 3011;

Best local similarity 81.4%; Pred. No. 1e-135;

Matches 320; Conservative 36; Mismatches 37; Indels 0; Gaps 0;

```
QY 1 MAAACGAGVFTGLTLSPYKYLARLWMLOYLTVREAHQWVPLNVRGGRDAI 60
Db 814 VAAACGAGVFTGLTLSPYKYLARLWMLOYLTVREAHQWVPLNVRGGRDAI 873
QY 61 ILTCAVHPELIDITKLLAFGLPLVLAQGITKVEYFVAQGLIRACMLVRKAGHY 120
Db 874 ILTCAVHPELIDITKLLAFGLPLVLAQGITKVEYFVAQGLIRACMLVRKAGHY 933
QY 121 VQAFMKLAATGTVYVVDHLTPLODMNAGRLDAVAVEPIFEDMEVKIITWADTAAC 180
Db 934 VQAFMKLAATGTVYVVDHLTPLODMNAGRLDAVAVEPIFEDMEVKIITWADTAAC 993
QY 181 GDIISGLPVSARRGRELILGPADNFEQGWELAPITAYSGQOTGLGCIITSLTGKDN 240
Db 994 GDIISGLPVSARRGRELILGPADNFEQGWELAPITAYSGQOTGLGCIITSLTGKDN 1053
QY 241 QVEGEVQIVSTATQTFPLATCINGVCMTVYHGAGTFTIASPKGPVLIQTYTNVDOLVGNPA 300
```

```
Db 1054 QVEGEVQIVSTATQTFPLATCINGVCMTVYHGAGTFTIASPKGPVLIQTYTNVDOLVGNPA 1113
QY 301 PGARSMTPTCTGSSDLYLVTRHADVI PVRRGDSRGSLSLSPRVSYLKSSGGPILCPG 360
Db 1114 PGARSMTPTCTGSSDLYLVTRHADVI PVRRGDSRGSLSLSPRVSYLKSSGGPILCPG 1173
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 393
Db 1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206
```

## RESULT 9

```
JCS620
genome polypeptide - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
Protein NS4; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Accession: JCS620
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A:Reference number: JCS620; MUID:9736593; PMID:9223423
A:Accession: JCS620
A:Molecule type: mRNA
A:Residues: 1-3014 <N>
A:Cross-references: GB:Y1184
A:Experimental source: genotype 5a, which predominates in South Africa
A:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
F:1.115/Product: capsid protein C #status predicted <CPC>
F:1.191/Product: envelope protein M #status predicted <EMP>
F:1.389/Product: major envelope protein E #status predicted <MEB>
F:1.390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:1.707-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1.1007-1616/Product: hepatitisin #status predicted <NS4>
F:1.130-1237/Product: nucleotide-binding motif A (P-loop)
F:1.1312-1317/Region: nucleotide-binding motif B
F:1.1316-1319/Region: DEXH motif
F:1.1617-1863/Product: nonstructural protein NS4 #status predicted <NS4>
F:1.1664-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2.015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:1.196,209,334,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23
```

Query Match 72.7%; Score 1493; DB 1; Length 3014;

Best local similarity 68.8%; Pred. No. 2.3e-114;

Matches 267; Conservative 56; Mismatches 65; Indels 0; Gaps 0;

```
QY 6 GCAVFTGLTLSPYKYLARLWMLOYLTVREAHQWVPLNVRGGRDAIILTC 65
Db 820 GCAVFTGLTLSPYKYLARLWMLOYLTVREAHQWVPLNVRGGRDAIILTC 879
QY 66 AVHPELIDITKLLAFGLPLVLAQGITKVEYFVAQGLIRACMLVRKAGHYQMAF 125
Db 880 AVHPELIDITKLLAFGLPLVLAQGITKVEYFVAQGLIRACMLVRKAGHYQMAF 939
QY 126 MKLAATGTVYVVDHLTPLODMNAGRLDAVAVEPIFEDMEVKIITWADTAACDITS 185
Db 940 MKLAATGTVYVVDHLTPLODMNAGRLDAVAVEPIFEDMEVKIITWADTAACDITS 999
QY 186 GDIISGLPVSARRGRELILGPADNFEQGWELAPITAYSGQOTGLGCIITSLTGKDN 245
Db 1000 GDIISGLPVSARRGRELILGPADNFEQGWELAPITAYSGQOTGLGCIITSLTGKDN 1059
QY 246 VQAFMKLAATGTVYVVDHLTPLODMNAGRLDAVAVEPIFEDMEVKIITWADTAAC 305
Db 1060 VQAFMKLAATGTVYVVDHLTPLODMNAGRLDAVAVEPIFEDMEVKIITWADTAAC 1119
QY 306 SMPPTCTGSSDLYLVTRHADVI PVRRGDSRGSLSLSPRVSYLKSSGGPILCPGSHAG 365
Db 1120 SMPPTCTGSSDLYLVTRHADVI PVRRGDSRGSLSLSPRVSYLKSSGGPILCPGSHAG 1179
```

366 IFRAAVCTRGVAKAVDFIPVESMETTR 393  
1180 VFRAAVCTRGVAKALFEVPEVNETTR 1207

RESULT 10

Q01303 genome polyprotein - hepatitis C virus (isolate HC-J6)

N:contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 17-Nov-2000

C:Accession: J01303

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kural, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.

J. Gen. Virol. 72, 2697-2704, 1991

A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human

A:Reference number: J01303; MUID:9204440; PMID:1658136

A:Accession: J01303

A:Molecule type: genomic RNA

A:Residues: 1-3033 <OKA>

A:Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651

A:Experimental source: isolate HC-J6 from a Japanese individual

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolyase; P-loop; polyprotein; serine proteinase; trans

F:115/191/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:330-733/Product: nonstructural protein NS1 #status predicted <NS1>

F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>

F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>

F:1316-1321/Region: nucleotide-binding motif B

F:1320-1323/Region: DEXH motif

F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>

F:1667-2017/Product: nonstructural protein NS4b #status predicted <N4B>

F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 68.3%; Score 1403; DB 1; Length 3033;

Best Local Similarity 64.9%; Pred. No. 6,2e-107;

Matches 252; Conservative 58; Mismatches 78; Indels 0; Gaps 0;

QY 6 GGAFFGIALTLTSPYKVLARLIMVLYITRVEAHLOVWIPPLNVRGGRDAIILTLTC 65  
DB 823 GAAVLVITITLTPGKTLISRFIMWLCTLAEMVQEMAPPOVNGRGCIIMVAV 882  
QY 66 AVHDELIFDITKLLAIFGLMVLQAGITVYFVRAQGLIRACMLVRAAGSHYVQAF 125  
DB 883 IFPGVVFEDITKMLAVLGPAYLLKGLITREPVFVAHALRMCITVVRHLAGRYVQVNL 942  
QY 126 MKLAALGTVYVHLTPLODMAHAGRLDVAVPEVFSMEVKIITWGADTAAGDIIS 185  
DB 943 IALGRMTGITTYDHLTPMSMAANGRLDVAVPEVFSMEVKIITWGADTAAGDIIS 1002  
QY 186 GLPVSARGRRIILGPADNEGGQWRLLAPITAYSOQTRGLLGCITSLTGRDKQVEGE 245  
DB 1003 GLPVSARLGRVILGPADGYTSKGMSLAPITAYAOQTRGLLGCITVIVSMTRGRDKQAGE 1062  
QY 246 VQVAVSTATQSLATCVNGVCTVFEHAGSKTLAGPRGPIITQVNTNDOLVGMQAPRGAR 305  
DB 1063 IQVLSVYQTFGLTISGVLMVTHGAGNKTLAGSGPVTQMTSSAEGDLVGMPSPGTK 1122  
QY 306 SMTPTCGSSDLYLVTRHADVIFVRRRGRSGSLSPRVSYLKSGSGPGLCPSGHANG 365  
DB 1123 SLRPTCGAVDYLVTVRMADVIPARRRGRGALLPRPLSTLKSSGGPVLCPRHANG 1182  
QY 366 IFRAAVCTRGVAKAVDFIPVESMETTR 393  
DB 1183 VFRAAVCTRGVAKALFEVPEVNETTR 1207

RESULT 11

GNWYJ8

genome polyprotein - hepatitis C virus (strain HC-J8)

N:contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 19-Jan-2001

C:Accession: A40250; P00397; P00559

R:Okamoto, H.; Kural, K.; Okada, S.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.;

Virol. 188, 331-341, 1992

A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to report

A:Reference number: A40250; MUID:9223032; PMID:1314459

A:Accession: A40250

A:Molecule type: genomic RNA

A:Residues: 1-3033 <OKA>

A:Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BA01761.1; PID:G221609

A:Experimental source: isolate E-b12

R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,

Biochem. Biophys. Res. Commun. 181, 279-285, 1991

A:Title: Distribution of plasmal HCV types in Japan.

A:Reference number: P00559; MUID:92068204; PMID:1720309

A:Accession: P00559

A:Molecule type: mRNA

A:Residues: 2678-2729 <KAT>

A:Cross-references: GB:D10562; GB:D090518; NID:G221523; PIDN:BA01418.1; PID:G221524

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolyase; nonstructural

F:115/191/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:330-733/Product: nonstructural protein NS1 #status predicted <NS1>

F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>

F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>

F:1316-1321/Region: nucleotide-binding motif A (P-loop)

F:1320-1323/Region: DEXH motif

F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>

F:1667-2017/Product: nonstructural protein NS4b #status predicted <N4B>

F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,223,239,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,235

Query Match 68.2%; Score 1401; DB 1; Length 3033;

Best Local Similarity 63.4%; Pred. No. 9e-107;

Matches 246; Conservative 67; Mismatches 75; Indels 0; Gaps 0;

QY 6 GGAFFGIALTLTSPYKVLARLIMVLYITRVEAHLOVWIPPLNVRGGRDAIILTLTC 65  
DB 823 GAAVLVITITLTPAKYKILSRVWMLSTVLAHQIQWVPLEVRGRDGIIMVAV 882  
QY 66 AVHDELIFDITKLLAIFGLMVLQAGITVYFVRAQGLIRACMLVRAAGSHYVQAF 125  
DB 883 ILHRLVFEVTKMLLALIGPAYLLKASLRLIPYVRAHALRVCITVLAARVYQVNL 942  
QY 126 MKLAALGTVYVHLTPLODMAHAGRLDVAVPEVFSMEVKIITWGADTAAGDIIS 185  
DB 943 IIRGRMTGITTYDHLTPMSMAANGRLDVAVPEVFSMEVKIITWGADTAAGDIIS 1002  
QY 186 GLPVSARGRRIILGPADNEGGQWRLLAPITAYSOQTRGLLGCITSLTGRDKQVEGE 245  
DB 1003 GLPVSARLGRVILGPADGYTSKGMKLLAPITAYTQTRGLLGCITVIVSMTRGRDKQAGE 1062  
QY 246 VQVAVSTATQSLATCVNGVCTVFEHAGSKTLAGPRGPIITQVNTNDOLVGMQAPRGAR 305  
DB 1063 IQVLSVYQTFGLTISGVLMVTHGAGNKTLAGSGPVTQMTSSAEGDLVGMPSPGTK 1122  
QY 306 SMTPTCGSSDLYLVTRHADVIFVRRRGRSGSLSPRVSYLKSGSGPGLCPSGHANG 365  
DB 1123 SLRPTCGAVDYLVTVRMADVIPARRRGRGALLPRPLSTLKSSGGPVLCPRHANG 1182  
QY 366 IFRAAVCTRGVAKAVDFIPVESMETTR 393  
DB 1183 VFRAAVCTRGVAKALFEVPEVNETTR 1207

Db 1123 SLDPCTCGAVDLVTRNADVIPRRKDRGALLSPRLSTLKSSGGPVLCSRGHANG 1182  
 QY 366 IPRAAVCTRGVAKAVDFIPVESMETMR 393  
 Db 1183 LPRAAVCAAGVAKSIDIFIPESLDVATR 1210

## RESULT 12

polyprotein - douroucouli hepatitis GB virus A  
 T08841  
 C/Species: douroucouli hepatitis GB virus A  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C/Accession: T08841  
 J. Gen. Virol. 79, 41-45, 1998  
 R. Ecker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08841  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-3005 <ERK>  
 A/Cross-references: EMBL:AF023425; NID:G2828599; PIDN:AA040501.1; PID:G2828600  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 19.9%; Score 408.5; DB 2; Length 3005;  
 Best Local Similarity 31.3%; Pred. No. 6.5e-25;  
 Matches 11; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

QY 54 RGGSD-----AIIITCAVHELIIFDIKLLAIFGLMVLQAGITKVPYFVRAQGLIRAC 109  
 Db 813 RGGSDVTVAVWAAAGIIFREVRCALTA-----LAALLDSIDVLETL-ILTA 864  
 QY 110 MLVKA-----AGHYVQMAFMKLAALTGYVYDHLTPLOMAHAGLRDAVAEPV 161  
 Db 865 QPARAARLDLSTFLGADLTRAFAVRRLERGGVTLFQHCQVSGXAAAILXDGVALPEV 924  
 QY 162 IFSMEKILTWGDTACGDIIISGLVSAARGREILG--PADNFGOGRLIAPITAY 219  
 Db 925 SVTARDCYVADARFALACGQREGLFVVAKEGVAVGVFSPRALPRGFVPAEPV- 993  
 QY 220 SQQTRGLIGCIITSLTGKXQVGEVQVSTATQSEFLATCVNGCVTFEGAGSKTLAG 279  
 Db 984 MQRLGFFSVYKTMGLGDERHSGIVLGTSTRMGTCVNGVMYTTFHGSNARTLAG 1043  
 QY 280 PKGPIITQMTYVDDLDVGMQAPPRARSMPTCTGSSDPLVYTRADVIPRRGDSRGL 339  
 Db 1044 PVGCVNCRWSPSDVAVYPLPSGASCIEPCCKCTQGVWCIRN--DALCHGRLSKVEL 1101  
 QY 340 LSPFVSYLKSSGSGPLCPGHAAGIFRAVCTRGV-----AKAVDFIPVES 387  
 Db 1102 DLPIEISDFRSSGSPILCDRGHYVGMV-VSLHKGKVGVRVYKMETLPKXS 1155

## RESULT 13

T08839  
 polyprotein - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: T08839  
 R. Ecker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <ERK>  
 A/Cross-references: EMBL:AF023424; NID:G2828597; PIDN:AA040501.1; PID:G2828598  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 16.7%; Score 342.5; DB 2; Length 2970;

Best Local Similarity 28.9%; Pred. No. 1.8e-19;  
 Matches 103; Conservative 56; Mismatches 133; Indels 65; Gaps 12;

QY 78 LLALFGP-----LMTIAGITKVPYFVRAQGLIRACMLVRAAGHYVQMAFPAKLA 129  
 Db 816 LVAAFWFVREIAAVCAVAFILFGFEDVDVILEVAVSSPVLVRLARLDSLVAAGDKLA 875  
 QY 130 ALTGYYVDHLTPLOD--MAAG-----LRDLAAVEPVIESDVEVITWGADTA 178  
 Db 876 T--TWLVEKRRKNCFLYAAAGOVTRTAQOLRWKGRALBPVAVHPEDCAVMDARTL 932  
 QY 179 ACGDIISGLPVSARREIILGPADNFGOGRL-----LAPITAVSQQTRGLIGCIIT 232  
 Db 933 SCGQSVHGRKPVVARRGDEVLLGVNGV--WELPPGFVTPAVVH--HKGGFPGVVK 987  
 QY 233 SITGRDKQVGEVQVSTATQSEFLATCVNGCVTFEGAGSKTLAGKPIITQMTYV 292  
 Db 988 SMTGWDETEHVNIVVLTGSTRSMGTCVNGVMYTTFHGSNARTLAQMGVNSRWMSAS 1047  
 QY 293 ODVGMQAPPGARSMPTCTGSSDPLVYTRADVIPVRRGDSRGLS----- 341  
 Db 1048 DVAVYPLPVAKCEPCKCPQGVWV-----RND--GALCHGTIGRTVELDL 1094  
 QY 342 PRPVSYLKSSGSGPLCPGHAAGIFRAVCTRG-----VAXAVDFIPVESMET 391  
 Db 1095 PAILCDPFGSSGSPILCDRGHYVGMV-LSVLHRSRVTVIRYKEMETLPRBAITHT 1150

## RESULT 14

H71426  
 hypothetical protein - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 A/Variety: Columbia  
 C/Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
 C/Accession: H71426  
 R. Ewan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; Giel  
 avenagh, T.; Hempel, S.; Kotter, P.; Eitlan, K.D.; Rieger, M.; Schaeffer, W.; Funk, B.  
 Nature 391, 485-488, 1998  
 A/Authors: Wellenreger, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdemanech,  
 etnot, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Beres, V.; Rechman, S.; Anst  
 C.; Chalmers, N.  
 A/Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal:  
 A/Reference number: A71400; MUID:98121113; PMID:9461215  
 A/Accession: H71426  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-692 <REV>  
 A/Cross-references: GB:Z97340; NID:G2244950; PID:G327492; PID:G2244965  
 C/Genetics:  
 A/Map position: 4COP9-4G3845

Query Match 5.5%; Score 112; DB 2; Length 692;  
 Best Local Similarity 24.1%; Pred. No. 0.3;  
 Matches 97; Conservative 45; Mismatches 135; Indels 126; Gaps 23;

QY 26 IARLIMLOGLITRFAHILQVMIPLVNRGGRDALITLCAVHPELIPITKULLAIFGP 85  
 Db 83 IADLAFGINVLMR-----QGNFPTSVVAGNSCEL-----KPEIIMDLTEILR--FLT 131  
 QY 86 IAWLQGITKVPY--FVRAQGLIRACMLVR--KAAAGHYVQMAFM----- 126  
 Db 132 LCMV---FSKRPVAVLESAGYHEDVLLQKPKAGVGHIMQRAFTIIRDTNSKILLR 188  
 QY 127 -----KLAALGTYYVDHLTPLODMAHAGLRDLAAVAEVPVIESDVEVITWGADTA 179  
 Db 189 GTSHIKDTLTAAGAVVPRFHSVLAHD--GGLSNLVLGY-----NH 226  
 QY 180 CGDIIISGLPVSARREIILGPADNFGOGRL-----LAPITAVSQQTRGLIGCIITSLGR 237  
 Db 227 CG--MVAAR-----WAKLSP-----CLTKXL----- 248  
 QY 238 DKQVGEVQVSTATQSEFLATCV-----NGVCWTVFHAGSKTLAGPKGP--I 284

```
DB 249 DENP-SFNVQIVGHSLSGGTASLTLYIREQKERASATCTFPAGTFLMINGSKGFI 307
QY 285 TQMTNVNDQDLY---GWQAPPGASMTPTCTGSSDLYLVTNHADVIP-VRRGDSRGSL 340
DB 308 TTIING--SDLVPTFSASSVDLSEVTSSSMSNDLRQVEHTRVLSVYRSATATIGSRL 365
QY 341 SPRPVSYLKSSGGPILCP--SGHVGIFRAAVCTRGVAKAYD 381
DB 366 PSIASAKAKVAGAGAILRPVSSGTQVAAPLVNGC--GKIKCID 406
```

## RESULT 15

VHMMH2

structural protein 2 precursor - hepatitis E virus (strain Burma)

C:Species: hepatitis E virus

C&gt;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 23-Jul-1999

C/Accession: C40778

R:Itam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.

Virology 185, 120-131, 1991

A:Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi

A:Reference number: A40778; MUID:92024067; PMID:1926770

A:Accession: C40778

A:Molecule type: genomic RNA

A:Residues: 1-660 &lt;TAM&gt;

A:Cross-references: GB:M73218; NID:g310023; PIDN:AAA45736.1; PID:g330026

A&gt;Note: the authors translated the codon CGC for residue 2 as Ala

C:Superfamily: hepatitis E virus structural protein 2

C:Keywords: structural protein

F:1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F:23-660/Product: structural protein 2 #status predicted &lt;SP2&gt;

## Query Match

Best Local Similarity 5.0%; Score 102.5; DB 1; Length 660;

Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

```
QY 76 TKLL--AIFGLMVLQAG-----ITKPYFVR--AQLIPACMLVRKAGHYQMA 124
DB 151 TNLVLAAPLSPPLPLQDGNTHIMATEASNYAQRVARATIRRLVPMVAVGVAISIS 210
QY 125 FMKLAALTGYVVDHLTPLODWAHAGLDLAVAVEPIFSDMEVKITWGADTAACDII 184
DB 211 FWPQTITPTTSV-----DNMSITDVRILVQPIASELYI----- 246
QY 185 SGLVSAARQRREILGPAD--NFEQGQRLAPI-TAYSQOTRGLL----GCITSLNG 236
DB 247 -----PSERLHYRNQGMRSVETSGVAEEATSGLVMLCIHGSIVNSYTN 290
QY 237 -----RDKXQVEGEVQVSTATQSFL 257
DB 291 TPYTGALGLIDFALELEFRMLTPGNTNTRVSRYSSTARHLRGRGADGTALTTTAATRFM 350
QY 258 A---TCVNGV---CMTVFH-----GAG-----SKTLGAPKG-PIT 285
DB 351 KDLYFTSTNGVGEIGRIALTFLNLADTLGLPTELLISSAGQLFYSRPVVSANGEPV 410
QY 286 QMTNTVDQDLYVQWAPPGASMTPTCTGSSDLYV--TRHADVIPVRRGDSRG-SLIS 341
DB 411 KLTYSVENA---QODKGIAPHDIDGESRVVIQDYDQHEQDRPTSPAPSRPFSYLR 466
QY 342 PRPVSYLK-----GSSGAPLCPGSHAVGIFRAAVCTRGVAKAYDFIPV 385
DB 467 ANDVLMWLSLVAARYDQSTYSSSTGCPYV--SDSVTLVNVATGAGAVARSLDWTKV 519
```

Search completed: May 6, 2004, 09:37:17  
Job time: 14.7992 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 8.20459 Seconds  
(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-11

Perfect score: 2053  
Sequence: 1 MAASCGAVFGLALITSP.....RGVAKAVDTFVPSMETTKR 393

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1967	95.8	3010	POLG_HCVUA
2	1950	95.0	3010	POLG_HCVUT
3	1935	94.3	3010	POLG_HCVTW
4	1888	92.0	3010	POLG_HCVBX
5	1764	85.9	3011	POLG_HCVL
6	1752	85.3	3011	POLG_HCVH
7	1403	68.3	3033	POLG_HCVU6
8	1401	68.2	3033	POLG_HCVU8
9	102.5	5.0	660	VST2_HEVU
10	102.5	5.0	660	VST2_HEVU
11	101.5	4.9	1780	POLG_MVEV
12	101	4.9	564	SR5C_ARATH
13	101	4.9	600	P37107 arabisdopsis
14	94	4.6	444	YAA7_XYLPA
15	93.5	4.6	1380	Q27675 leishmania
16	93	4.5	434	TOLB_CHLLE
17	92.5	4.5	706	TRFE_HORSE
18	92	4.5	444	V327_XYLPT
19	92	4.5	659	VST2_HEVME
20	92	4.5	3414	POLG_TBEVH
21	91.5	4.5	401	PILC_PSEBU
22	91.5	4.5	485	VST2_HEVH
23	91.5	4.5	660	VST2_HEVH
24	91.5	4.5	3412	POLG_TBEVS
25	90.5	4.4	403	KHPT_ECOLI
26	90.5	4.4	3414	POLG_TBEVH
27	88.5	4.4	961	ARCW_YERPE
28	88.5	4.3	355	YORG_MEIMA
29	88.5	4.3	3432	POLG_JAEVA
30	88	4.3	441	KDHP_MESCR
31	87.5	4.3	3432	POLG_JAEV1
32	87.5	4.3	3432	POLG_JAEVU
33	86.5	4.2	347	MDHM_EUCGU

34	86.5	4.2	470	1	NRAM_IATRM	P05803 influenza a
35	86	4.2	338	1	GALE_MEICO	O05026 neisseria g
36	86	4.2	433	1	DCUA_MOLSU	O34245 wolfeella s
37	86	4.2	790	1	RIRI_HSVB	P28846 equine herp
38	85.5	4.2	355	1	YP71_MYCRU	O50650 mycobacteri
39	85	4.1	470	1	NRAM_IATRA	P03472 influenza a
40	85	4.1	730	1	HELS_MEIMA	O8p2r7 methanocarc
41	85	4.1	654	1	PMP2_SCHPO	O9c1x1 schizosacch
42	85	4.1	313	1	CIR3_RAT	O88278 rattus norv
43	84.5	4.1	453	1	NRAQ_BACSU	O07553 bacillus su
44	84.5	4.1	1705	1	PIPV_MOUSE	P70289 mus musculu
45	84	4.1	305	1	OPFC_BACSU	P24139 bacillus su

## ALIGNMENTS

RESULT 1  
POLG\_HCVUA  
ID POLG\_HCVUA STANDARD: PRT: 3010 AA..

AC P26662; 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
DE Hepatitis C virus (Isolate Japanese) (HCV).  
OS Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepatitis virus.  
NCBI\_TaxID=11116;  
OX NCBI  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91088550; PubMed=2175903;  
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;  
RT "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis".  
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528 (1990).  
RN [2]  
RX MEDLINE=91192160; PubMed=1849488;  
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K., Ohkoshi S., Shimotohno K.;  
RT "Molecular structure of the Japanese hepatitis C viral genome.";  
RL FBS Lett. 280:325-328 (1991).  
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).  
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipid protein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and RNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: D90208; BAA14233.1; -

Length 3010;

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL contribution to the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D11168; BA001943.1; --  
 DR PIR; A45573; A45573.  
 DR MEROPS; S29.001; --  
 DR MEROPS; U39.001; --  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002552; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam; PR01543; HCV\_capsid; 1.  
 DR Pfam; PR01542; HCV\_core; 1.  
 DR Pfam; PR01539; HCV\_env; 1.  
 DR Pfam; PR01560; HCV\_NS1; 1.  
 DR Pfam; PR01538; HCV\_NS2; 1.  
 DR Pfam; PR02907; HCV\_NS3; 1.  
 DR Pfam; PR01006; HCV\_NS4a; 1.  
 DR Pfam; PR01001; HCV\_NS4b; 1.  
 DR Pfam; PR01506; HCV\_NS5a; 1.  
 DR Pfam; PR00271; Helicase\_C; 1.  
 DR Pfam; PR00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KM Core protein; Coat protein; Helicase; ATP-binding;  
 KM Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.  
 FT INT\_MET 1  
 FT CHAIN 1  
 FT CHAIN 115  
 FT CHAIN 192  
 FT CHAIN 383  
 FT CHAIN 729  
 FT CHAIN 1006  
 FT CHAIN 1007  
 FT CHAIN 1615  
 FT CHAIN 1616  
 FT CHAIN 1863  
 FT CHAIN 2013  
 FT CHAIN 2014  
 FT CHAIN 347  
 FT TRANSMEM 1083  
 FT ACT\_SITE 1107  
 FT ACT\_SITE 1107  
 FT ACT\_SITE 1165  
 FT ACT\_SITE 1230  
 FT NP\_BIND 1237  
 FT SITE 1316  
 FT CARBOHYD 1316  
 FT CARBOHYD 196  
 FT CARBOHYD 209  
 FT CARBOHYD 234  
 FT CARBOHYD 250  
 FT CARBOHYD 305  
 FT CARBOHYD 317  
 FT CARBOHYD 417  
 FT CARBOHYD 423  
 FT CARBOHYD 430  
 FT CARBOHYD 448  
 FT CARBOHYD 532  
 FT CARBOHYD 540  
 FT CARBOHYD 556  
 FT CARBOHYD 576

FT CARBOHYD 623  
 FT CARBOHYD 645  
 FT CARBOHYD 2041  
 FT CARBOHYD 2077  
 FT CARBOHYD 2240  
 FT CARBOHYD 2529  
 FT CARBOHYD 2788  
 FT CARBOHYD 3010  
 FT CARBOHYD 326573  
 FT CARBOHYD 9441C77435D642B CRC64;  
 FT CARBOHYD 95.0%; Score 1950; DB 1; Length 3010;  
 FT CARBOHYD Best Local Similarity 93.6%; Pred. No. 4.2e-150;  
 FT CARBOHYD Matches 368; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKYLARLWLYLITREVAHLQVWIPPLNVRGRDAI 60  
 DB 814 MAASCGAVFGLALTLSPYKYLARLWLYLITREVAHLQVWIPPLNVRGRDAI 873  
 QY 61 ILTCAVPELIFDITKLLAIFGPIMLQAGITVPEVFAQGLIRACMLVRKAGGHY 120  
 DB 874 ILTCAVPELIFDITKLLAIFGPIMLQAGITVPEVFAQGLIRACMLVRKAGGHY 933  
 QY 121 VQMAFMKLAALGTIVYDHLPLQDMAHAGRLDAVAEPIVFSMEVKIITWGAADTAAC 180  
 DB 934 VQMAFMKLAALGTIVYDHLPLQDMAHAGRLDAVAEPIVFSMEVKIITWGAADTAAC 993  
 QY 181 GDIISGLPVSARREIILGPADNFEQGMRLAPITVSCQTRLLGCIITSLTGRKXN 240  
 DB 994 GDIISGLPVSARREIILGPADNFEQGMRLAPITVSCQTRLLGCIITSLTGRKXN 1053  
 QY 241 QVEGEVQVSTATQSEFLATCVNGVCMTPFHGAGSTLPGKPIQMTNTVNDQDLVQQA 300  
 DB 1054 QVEGEVQVSTATQSEFLATCVNGVCMTPFHGAGSTLPGKPIQMTNTVNDQDLVQQA 1113  
 QY 301 PPGARSMTPECTGSSDLYLVTRHADVIPVRARGDSRGLISPRVSYLKSGSGGFLCPS 360  
 DB 1114 PPGARSMTPECTGSSDLYLVTRHADVIPVRARGDSRGLISPRVSYLKSGSGGFLCPS 1173  
 QY 361 GHAIVGFRAAVCTRGAAVDFIPVESMETTR 393  
 DB 1174 GHAIVGFRAAVCTRGAAVDFIPVESMETTR 1206

## RESULT 3

POLG\_HCVTM  
 ID POLG\_HCVTM STANDARD; PRT; 3010 AA.

DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polypeptide [contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 DE Hepatitis C virus (isolate Taiwan) (HCV).  
 OS Hepatitis C virus (isolate Taiwan) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=31645;  
 RN [1]  
 RP MEDLINE=92230206; PubMed=1314449;  
 RX Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
 RA "The Taiwan hepatitis C virus genome: sequence determination and  
 RT mapping the 5' terminus of viral genome and antigenomic RNA.";  
 RL Virology 188:102-113(1992).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.



FT	CARBOHYD	233	233	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT <td>CARBOHYD</td> <td>234</td> <td>234</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	234	234	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>250</td> <td>250</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	250	250	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>305</td> <td>305</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	305	305	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>417</td> <td>417</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	417	417	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>423</td> <td>423</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	423	423	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>430</td> <td>430</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	430	430	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>448</td> <td>448</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	448	448	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>532</td> <td>532</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	532	532	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>540</td> <td>540</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	540	540	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>556</td> <td>556</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	556	556	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>576</td> <td>576</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	576	576	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>623</td> <td>623</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	623	623	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>645</td> <td>645</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	645	645	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>2041</td> <td>2041</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	2041	2041	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>2077</td> <td>2077</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	2077	2077	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>2240</td> <td>2240</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	2240	2240	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>2529</td> <td>2529</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	2529	2529	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT <td>CARBOHYD</td> <td>2788</td> <td>2788</td> <td>N-LINKED<td>(GLCNAC .)<td>(POTENTIAL)</td></td></td>	CARBOHYD	2788	2788	N-LINKED <td>(GLCNAC .)<td>(POTENTIAL)</td></td>	(GLCNAC .) <td>(POTENTIAL)</td>	(POTENTIAL)
SQ	SEQUENCE	3010 AA;	327047 MW;	AAD267D5CDEF2215 CRC64;		
Query Match 94.3%; Score 1935; DB 1; Length 3010;						
Best Local Similarity 91.9%; Pred. No. 7e-149;						
Matches 361; Conservative 17; Mismatches 15; Indels 0; Gaps 0						
QY	1	MAASCGAVFGLALTLTSPYKYLARLIMWLQYLITRVEAHLOVMIDPLNVRGRDAI	60			
DB	814	MAASCGAVFGLALTLTSPYKYLARLIMWLQYLITRVEAHLOVMIDPLNVRGRDAI	873			
QY	61	ILITCAVHELFDITTKLLAFGLIWLQAGIKVYFFRAGGLRACMLVRKAAGHY	120			
DB	874	ILITCAVHELFDITTKLLAFGLIWLQAGIKVYFFRAGGLRACMLVRKAAGHY	933			
QY	121	VQMAFMKLAALIGTVYVYDHLPTLPDMAGLRDLAVAEVPIISDMEVKITMGADTAAC	180			
DB	934	VQMAFMKLAALIGTVYVYDHLPTLPDMAGLRDLAVAEVPIISDMEVKITMGADTAAC	993			
QY	181	GDIIISLPVSARSGEILIGPADNBEQGRMLAPITANSQOQTRGLCITISLSTRDN	240			
DB	994	GDIIISLPVSARSGEILIGPADNBEQGRMLAPITANSQOQTRGLCITISLSTRDN	1053			
QY	241	QVEGEVQVNSTATOSFLATCNAGVCMVFFGAGSKTLAGKGPITQMYTNVDDLVGMQA	300			
DB	1054	QVEGEVQVNSTATOSFLATCNAGVCMVFFGAGSKTLAGKGPITQMYTNVDDLVGMQA	1111			
QY	301	PGASMTPTCTGSSDLYLTRADVIPRRGRSGSLISPPVSYLKGSSGGLLCP	360			
DB	1114	PGASMTPTCTGSSDLYLTRADVIPRRGRSGSLISPPVSYLKGSSGGLLCP	1172			
QY	361	GHAVGIFRAVCTRGVAKADPIVESEMTMR	393			
DB	1174	GHAVGIFRAVCTRGVAKADPIVESEMTMR	1206			
RESULT 4						
POLG HCVR						
AC	ID_POLG_HCVR	STANDARD;	PRT: 3010 AA.			
DT	01-AUG-1992	(Rel. 23, Created)				
DT	01-AUG-1992	(Rel. 23, Last sequence update)				
DT	10-OCT-2003	(Rel. 42, Last annotation update)				
DE	Genome polypeptide (Contains: Capsid protein C (Core protein) (p22);					
DE	Envelope glycoprotein E1 (GP33) (GP35); Envelope glycoprotein E2					
DE	(GP66) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (p21)					
DE	(EC 3.4.22.-); Protease/helicase NS3 (p70) (Hepatitisin)					
DE	(EC 3.4.21.98); Nonstructural protein NS4A (p4); Nonstructural protein					
DE	NS4B (p27); Nonstructural protein NS5A (p56); Nonstructural protein					
DE	NS5B (p66) (p70) (RNA-directed RNA polymerase) (EC 2.7.7.48)1.					
OS	Hepatitis C virus (isolate BX) (HCV).					
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;					
OC	Hepadnavirus.					
OX	NCBI_Taxid=11105;					
ON	[1]					



```

FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT STRAND 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT HELIX 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1186
FT TURN 1187 1188
FT STRAND 1189 1197
FT HELIX 1198 1202
FT TURN 1203 1204
FT STRAND 1203 1204
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDF9C CRC64;

Query Match 92.0%; Score 1888; DB 1; Length 3010;
Best local similarity 90.8%; Pred. No. 4,7e-145;
Matches 357; Conservative 13; Mismatches 23; Indels 0; Gaps 0;

QY 1 MAASCGAVPTGLALTLSPYKYLARLIMLQYLITRVEAHLQVWIPLVNRRGSDAI 60
DB 814 MAASCGAVPTGLALTLSPYKYLARLIMLQYLITRVEAHLQVWIPLVNRRGSDAI 873
QY 61 ILTLCANVPELITPTLLALITGLMVLQAGITKVPYFRAQGLIPACMLVKKAGCHY 120
DB 874 ILTLCANVPELITPTLLALITGLMVLQAGITKVPYFRAQGLIPACMLVKKAGCHY 933
QY 121 VQNAFMKLAALGTGYVDHLLPLQDMANAGLRDLAVAVEPITSDEVKIITWGADTAAC 180
DB 934 VQNAFMKLAALGTGYVDHLLPLQDMANAGLRDLAVAVEPITSDEVKIITWGADTAAC 993
QY 181 GDIISGLPVARRRRELLIPADNFEQSGRIILAPITAYSCQTRGLIGCTITSLTRDCK 240
DB 994 GDILGLPVARRRRELLIPADNFEQSGRIILAPITAYSCQTRGLIGCTITSLTRDCK 1053
QY 241 QVEGEVQVSTATQSFATCNGVCWTFVFGAGSKTLTAGKGPITQWYTNVDQDLVGMOA 300
DB 1054 QVEGEVQVSTATQSFATCNGVCWTFVFGAGSKTLTAGKGPITQWYTNVDQDLVGMR 1113
QY 301 PPGARSMTPTCGSSDLVLTTRADVTPVRRRSGSLISPPVSYLKSSGGPILCPG 360
DB 1114 PPGARSMTPTCGSSDLVLTTRADVTPVRRRSGSLISPPVSYLKSSGGPILCPG 1173
QY 361 GHAVGIFRAVCTRGVAKAVDFIPVESMETTMR 393
DB 1174 GHAVGIFRAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 5
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26654;
ID ID_P001;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (hepativirin)
DE (EC 3.4.21.38); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
DE Hepatitis C virus (isolate 1) (HCV).
OS Hepatitis C virus positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_Taxid=11104;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.O., Weiner A.O.,
RA Bradley D.W., Kuo G., Houghton M.;
RA "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication. NS3
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
CC {RNA} (N).
CC -1- SUMMARY: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-ch).
CC -----
DB EMBL: M62321; AAA45676.1; -.
DB PIR: A39166; GMYC3.
DB PDB: 1AIV; 16-FEB-99.
DB PDB: 1HEI; 25-NOV-98.
DB MEROPS: S29.001; -.
DB MEROPS: U39.001; -.
DR Interpro: IPR009003; Cys_Ser_trypsin.
DR Interpro: IPR001410; DEAD.
DR Interpro: IPR002822; HCV_capsid.
DR Interpro: IPR002821; HCV_core.
DR Interpro: IPR002519; HCV_env.
DR Interpro: IPR002531; HCV_NS1.
DR Interpro: IPR002518; HCV_NS2.
DR Interpro: IPR000745; HCV_NS4A.
DR Interpro: IPR001490; HCV_NS4B.
DR Interpro: IPR002868; HCV_NS5A.
DR Interpro: IPR002166; HCV_RdRp.
DR Interpro: IPR001650; Helicase_C.
DR Interpro: IPR004109; Peptidase_C29.
DR Interpro: IPR007094; RNA_pol_DS_PS.
DR Pfam: PF01543; HCV_capsid.1.
DR Pfam: PF01542; HCV_core.1.
DR Pfam: PF01539; HCV_env.1.
DR Pfam: PF01560; HCV_NS1.1.
DR Pfam: PF01538; HCV_NS2.1.
DR Pfam: PF02907; HCV_NS3.1.
DR Pfam: PF01006; HCV_NS4A.1.
DR Pfam: PF01001; HCV_NS4B.1.
DR Pfam: PF01506; HCV_NS5A.1.
DR Pfam: PF00271; Helicase_C.1.
DR Pfam: PF00998; Viral_RdRp.1.
DR ProDom: PD186052; HCV_NS1.1.
DR SMART: SM00487; DEXDC.1.
KM Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Helicase; ATP-binding;
KM Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KM 3D-structure.
FT INIT_MBT 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006

```

REMOVED FROM CAPSID PROTEIN C BY THE  
 CELLULAR AMINOPEPTIDASE.  
 CAPSID PROTEIN C (POTENTIAL).  
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT TRANSMEM 347 369 POTENTIAL.  
 FT ACT SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 327157 MW; 65F8C9447FCBAP9 CMC64;

Query Match 85.9%; Score 1764; DB 1; Length 3011;  
 Best Local Similarity 81.9%; Pred. No. 5.6e-135;  
 Matches 322; Conservative 34; Mismatches 37; Indels 0; Gaps 0;

1 MASCGGAVFGLALLSPYKYLARLWMLOYLTREAHQVWIPINRGSDAI 60  
 814 VAASCGGVAVGLALTLPYKRYISWCLMLOYLTREAHQVWIPINRGSDAV 873  
 61 ILATCAHPELIPITKLLAIFGPMVLOAGITKVFYFAOGLIRACMLVKKAGGHY 120  
 874 ILLMCAHPELIPITKLLAIFGPMVLOAGITKVFYFAOGLIRACMLVKKAGGHY 933  
 121 VQNAFMKLAITGYVVDHLTPLODAHAGIRDLAVAVERYISDMEVKITMGADTAAC 180  
 934 VQVWIKLGAITGYVVDHLTPLODAHAGIRDLAVAVERYISDMEVKITMGADTAAC 993  
 181 GDIISGLPVASAREGELLGPDNFGQWRLLAPITAAQOQRGLGCIITSLTGRDX 240  
 994 GDIISGLPVASAREGELLGPDNFGQWRLLAPITAAQOQRGLGCIITSLTGRDX 1053  
 241 QVEGEVAVSTATQSPATCVNGVCMVTFHAGSKTLAGKPGITOMTNTVNDOLVGMQA 300  
 1054 QVEGEVAVSTATQSPATCVNGVCMVTFHAGSKTLAGKPGITOMTNTVNDOLVGMQA 1113  
 301 PPGASMTPTCTGSSD.YIVTRHADV.PVRRGDSRSLSPPRVSLTKSSGGPILCP 360  
 1114 PPGASMTPTCTGSSD.YIVTRHADV.PVRRGDSRSLSPPRVSLTKSSGGPILCP 1173

361 GHAVGIFRAAVCTGKAVKAVDFIPVESMETTM 393  
 1174 GHAVGIFRAAVCTGKAVKAVDFIPVESMETTM 1206

RESULT 6  
 POLG HCVA STANDARD; PRT; 3011 AA.  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polypeptide (Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NCBI\_Taxid=1108;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=92052256; PubMed=1658800;  
 Inchauste G., Zebadee S., Lee D.H.H., Sugtani M., Nasoff M.,  
 Prince A.M.;  
 "Genomic structure of the human prototype strain H of hepatitis C  
 virus: comparison with American and Japanese isolates.";  
 Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 [2]  
 X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 MEDLINE=9731322; PubMed=9487654;  
 Yao N., Heeson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 "Structure of the hepatitis C virus RNA helicase domain.";  
 Nat. Struct. Biol. 4:463-467(1997).  
 [3]  
 X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 MEDLINE=98154321; PubMed=9493270;  
 Kim J.L., Morenssteern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 Murcko M.A., Lin C., Caron P.R.;  
 "Hepatitis C virus NS3 RNA helicase domain with a bound  
 oligonucleotide: the crystal structure provides insights into the mode  
 of unwinding.";  
 Structure 6:89-100(1998).  
 RL Structure 6:89-100(1998).  
 RT of unwinding.";  
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position. Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND NS5A.  
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M67463; AAA45534.1; -.  
 CC PIR; A36814; GNVVCH.  
 CC PDB; 1HRI; 25-NOV-98.  
 CC PDB; 1AIV; 16-FEB-99.  
 CC PDB; 1AIR; 17-JUN-98.  
 CC MEROPS; S29.001; -.  
 CC TRANSFAC; T04155; -.  
 CC InterPro; IPR009003; Cys\_Set\_trypsin.  
 CC InterPro; IPR001410; DED.  
 CC InterPro; IPR002522; HCV\_capsid.

Query Match	Best Local Similarity	85.3%; Score 1752; DB 1; Length 3011;
Matches 320; Conservative	36; Mismatches 37; Indels 0; Gaps 0;	
1 MAASCGAVFGLALLTLSPYKVLARIWLOYLITRREAHLOVWIPPLNVRGSGDAI 60	814 VNASCGVVLVGMALTLSPYKXYSICMMWLOFLTRVEAQLHVVVPLNVRGSGDAV 873	
61 ILLICAVHPEILIPITKLLAIFGLMVLQAGITKVPYFRAQGLIRACMLVKKAAGHY 120	874 ILLICVHPEILIPITKLLAIFGLMVLQAGITKVPYFRAQGLIRACMLVKKAAGHY 933	
121 VQAFMKLALGTGYVDHLPLDQMAHAGLRDLAAVEEVIISDMEVKIIITGADTAAC 180	934 VQAFMKLALGTGYVDHLPLDQMAHAGLRDLAAVEEVIISDMEVKIIITGADTAAC 993	
181 GDIISGLPVASARRGELLGPADNFEQGMRLIAPITAYSQOTRGLIGCIITSLTGRDN 240	994 GDIISGLPVASARRGELLGPADNFEQGMRLIAPITAYSQOTRGLIGCIITSLTGRDN 1053	
241 QVEEEVQVVSARQSLFATGVNGVCMVEFHGAGSKTLAGKGPITOMYATVDDOLVGMQA 300	1054 QVEEEVQVVSARQSLFATGVNGVCMVEFHGAGSKTLAGKGPITOMYATVDDOLVGMQA 1113	
301 PPGARRMPTCTCGSSDLVLYTRHADVYVVRARRDGRSLSPRPVSYLKSSGGPILLCP 360		

```

Db      1114 POGSRSLPTCTGSSDLYLVTRHADVTPVRRRGSRSLSPRISLYKSSGGPLLCPT 1173
OY      361 GHAAGIFRAVCTGCAKAVADPIFVSEMETTMR 393
Db      1174 GHAAGIFRAVCTGCAKAVADPIFVSEMETTMR 1206

RESULT 7
POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P26660;
DT 01-AUG-1992 (Ref. 23, Created)
DT 28-FEB-2003 (Ref. 41, Last sequence update)
DE Genome polyprotein (Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxId=11113;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Iizuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704 (1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position. Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein and envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D00944; BA00792.1; -
DR PIR; J01303; J01303.
DR HSSP; P27958; IHEI.
DR MEROPS; S29.001; -
DR MEROPS; U39.001; -
DR InterPro; IPR009003; Cys_ser_lyspsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4A.
DR InterPro; IPR001490; HCV_NS4B.
DR InterPro; IPR002868; HCV_NS5A.
DR InterPro; IPR002166; HCV_RdRP.

```

```

DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007095; RNA_pol_PS.
DR InterPro; IPR007094; RNA_pol_Psivir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KT Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
FT INIT_MET 1
FT CHAIN 1
FT CHAIN 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT ACT_SITE 1230 1241
FT SITE 1320 1323
FT SITE 1324 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 329165 MW; F957F5CIA273B59E CRC64;

```

Query Match 68.3%; Score 1403; DB 1; Length 3033;  
 Best local similarity 64.9%; Pred. No. 1.3e-105;  
 Matches 252; Conservative 58; Mismatches 78; Indels 0; Gaps 0;

```

OY      6 GGAVFGLALTLSPYKYLRLRWLYLITRVEAHQVIRPLNRRGGDAIILLTC 65
Db      823 GAALVLTILFTLTPGYKTLISRLFWVCYLLTLLEAMQCEAPMPQVRGSGDITIMAV 882
OY      66 AVHPELIPDITKLLAIFGLPVLVLAQGITKVEFYFAOGLIRACMLVRLAAGHYVQAF 125
Db      883 IFPGVVPDITKMLAVALGPAVLKGLATRVYFYFAHLLMLCMVMVRLNAGRVQVYL 942
OY      126 KTLAALTGYYVDHLTPLODMAHAGRLDAVAVEPIYSDEWVKITTCAGTLAACGDITS 185
Db      943 LALGRWTGYIVDHLTPMSDMANGLRLDAVAVEPIIFSPMEKVIIVGAETPAAGDILH 1002

```



QY 186 GLPVASARRGEILLGPADNFEQGMRLAITYSCOTRLGCTITSLTGDRXNOVEGE 245  
 DB 1003 GLPVASARRGEILLGPADNFEQGMRLAITYSCOTRLGCTITSLTGDRXNOVEGE 1062  
 QY 246 VQVSTSTQSEFLATCVGVCMTVFHAGSGSKTLGPKKPIQWNTVNDODLVGQAPPGAR 305  
 DB 1063 VQVSTSTQSEFLATCVGVCMTVFHAGSGSKTLGPKKPIQWNTVNDODLVGQAPPGAR 1122  
 QY 306 SMTPTCTGSSDLVYVTHAAVPIVRRRGRSGSLSPRPVYVKGSGGPELCPGSHAVG 365  
 DB 1123 SMTPTCTGSSDLVYVTHAAVPIVRRRGRSGSLSPRPVYVKGSGGPELCPGSHAVG 1182  
 QY 366 IFPAVCTRGVAKAVDPVPEVSMETMR 393  
 DB 1183 IFPAVCTRGVAKAVDPVPEVSMETMR 1210

RESULT 8  
 POLG\_HCVJ8 STANDARD; PRT; 3033 AA.

AC P26661; 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 OC Viruses; ssRNA, positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 CX NCBI\_TaxID=11115;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9223023; PubMed=1314459;  
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Iizuka H., Tanaka T.,  
 RA Fukuda S., Tenda F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes.";  
 RT Virology 188:331-341(1992).  
 RL CC  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D10988; BAA01761.1; -;  
 DR PIR; A40250; GNVJ8.  
 DR HSSP; P27858; IHEI.  
 DR MEROPS; S29.001; -;  
 DR MEROPS; U39.001; -;  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RBP.  
 DR InterPro; IPR004109; peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVLR.  
 DR Pfam; PF01542; HCV\_capsid; 1.  
 DR Pfam; PF01543; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; Viral\_RBP; 1.  
 DR Pfam; PF01862; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT ACT\_SITE 1234 1241  
 FT NP\_SIND 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 289 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 FT CARBOHYD 3033 3033  
 FT SEQUENCE 33017 MW; 1A173E7E3381FD1A CR64;

Query Match 68.2%; Score 1401; DB 1; Length 3033;  
 Best Local Similarity 63.4%; Pred. No. 1.9e-105;  
 Matches 246; Conservative 67; Mismatches 75; Indels 0; Gaps 0;

QY 6 GGNVFGALTLSPFYVTLARLIWLOQYITRVEALQWVIRPLNVRGRDAIILITC 65  
 DB 823 GLAIVVITISFTLPAYKILSRGVMWLSYVLVAEAOIQQVBPPLVGRGDIWAV 882







Query March 1987 4.9% Score 101.5; DB 1; Length 1780;  
Best Local Similarity 19.8%; Pred. No. 4.4;  
Matches 81; Conservative 49; Mismatches 136; Indels 143; Gaps 16

QY 7 GAVFGLALLTLSPYKVLRLIWLQYLITRVAHLQWVIPPINVRGRDAILTLTCA 66  
Db 1340 GAVLIGLALITSTGYSPPTMA-----AGLMICNP--NKKRGWPATEVLTAA 1382  
QY 67 VHEPLIFDTLLLAIFGRLWLQNGITKVFYFAVAGGLIARCMVYKRAAGHYVQMAFM 126  
Db 1383 -----VGMEFAIVGGLAELDIDMSVFP-----TIAGLMVSYVTSKATDWMLE 1422  
QY 127 KL-----AALTCT-----YYDHLTPLOMAHAGRLDAVAEPYIF 163  
Db 1428 RAADVSWAGAAITSTSERLDVQDDDDGPHLNDPGVPKIKWV---LRMTCLSV----- 1472  
QY 164 SDMEKILITWGAADTACGDIISGLFVSARRGHEILLSPADNFEQGG-----WRLLAPT 217  
Db 1480 ----AAITPMWALPSPAFGWYLT--LKTKRGGVFWMDTSPKXYPRGDTTPGYRITMA-- 1530  
QY 218 AYSQOTRELLGCIITSLTGRDQNVGEVQVAVSTATOSPILATCNAGVCTVPEAGSKTL 277  
Db 1531 -----RILG-----RYQAGVGVMHSGVFHITLHTTRGAAI 1562  
QY 278 AGPKPITQMTYNTVDQDLVWQAP-----PGARSMTPTCCSSDLY 318  
Db 1562 MSGEGRLTPYGVGNVEDRVTYGGPKMLDQKMGVVDVQWIVPEPGPAINVQT---KPGI 1614  
QY 319 LVTHADVIVPYRRRDSRGSLSLPPPVSLTKSSSGGELLCPSGHAVGIF 367  
Db 1619 EKTALGEI-----GAVSLDYPI---GTSGSPVNSVGEITIGLY 1653

RESULT 12  
SRSC\_ARATH STANDARD; PRT; 564 AA.

ID \_SRSC\_ARATH  
AC P371077.082570; Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Signal recognition particle 54 kDa protein, chloroplast precursor  
DE (SRP54) (54 chloroplast protein) (54CP) (FFC).  
GN FFC OR ATSG03940 OR P8F6\_150.  
GN GN  
OS Arabidopsis thaliana (Mouse-ear cress).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.  
OC NCBI\_TaxId=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RC MEDLINE=94012817; PubMed=8408079;  
RA Franklin A.E.; Hoffman N.E.;  
RA "Characterization of a chloroplast homologue of the 54-kDa subunit of  
RT the signal recognition particle."  
RT J. Biol. Chem. 268:22175-22180(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;  
RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc,  
RT encoding the 54 kDa subunit of chloroplast signal recognition  
RT particle."  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RC MEDLINE=21016721; PubMed=11130714;  
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
RA Kohira M., Matsunoto M., Matsuno A., Muraki A., Nakayama S.,  
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,  
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
RA Huang E., Spielger L., Gnoj V., O'Shaughnessy A., Preston R.,

```

Ra Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
Ra Sconeck T., Pepin K., Spieth U., Seikton M., Armstrong J., Becker M.,
Ra Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Danne M.,
Ra Du H., Edwards J., Fryman J., Hansen B., Lamar E., Latrille P.,
Ra Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromatt C.,
Ra Wagner-McPherson C., Woliam A., Yoakam B., Bell W., Dedina N.,
Ra Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
Ra Kirchhoff K., Toeh K., King L., Bahret A., Miller B., Marxa M.A.,
Ra Mattianssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
Ra Voickert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
Ra Enrican K.D., Terryn N., Hartley N., Bent E., Johnson S.,
Ra Langham S.-A., McCullash B., Robben U., Grymopre B., Zimmermann W.,
Ra Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
Ra van Stevenen M., Dirkse W., Woolman P., Klein Lankhorst R.,
Ra Weltzeneger T., Boche G., Rose M., Haut F., Bernier S., Hempel S.,
Ra Feldpausch M., Lambirth S., Villarroel R., Gielen U., Adiles W.,
Ra Berts O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
Ra Schueller C., Zaccaria P., Mexes H.-W., Beyan M., Franz P.,
Ra "Sequence and analysis of chromosome 5 of the plant Arabidopsis
Rt thaliana."
Rl Nature 408:823-826(2000).
Cc -I- FUNCTION: May target chloroplast proteins to either the thylakoid
Cc or envelope membranes.
Cc -I- SUBCELLULAR LOCATION: Chloroplast stroma.
Cc -I- TISSUE SPECIFICITY: Most abundant in green shoot tissue and
Cc lower levels seen in roots and etiolated buds.
Cc -I- SIMILARITY: Belongs to the GTP-binding SRP family.
Cc -----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL outstation at
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc EMBL; Z21970; CAAT9981.1; -.
Cc EMBL; AF092168; AACG4139.1; -.
Cc EMBL; AL162873; CAB85514.1; -.
Cc PIR; S36637; S36637.
Cc HSSP; O07347; 1PFH.
Cc InterPro; IPRO03593; AAA ATPase.
Cc InterPro; IPRO00897; SRP54.
Cc InterPro; IPRO004125; SRP54.SPB.
Cc InterPro; IPRO04780; SRP_sub.
Cc Pfam; PF00448; SRP54_1.
Cc Pfam; PF02881; SRP54_N_1.
Cc Pfam; PF02978; SRP_SPB; 1.
Cc Prodom; PD00819; SRP54_1.
Cc SMART; SMO0382; AAA; 1.
Cc TRIFAMS; TRIG00959; ffh; 1.
Cc PROSITE; PS00300; SRP54; 1.
Cc Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;
Cc Transil peptide.
Cc KW TRANSIT 1 75 CHLOROPLAST.
Cc FT FT 76 564 SIGNAL RECOGNITION PARTICLE 54 Xda
Cc FT FT 75 PROTEIN.
Cc FT DOMAIN 76 370 G-DOMAIN.
Cc FT NP_BIND 371 564 M-DOMAIN.
Cc FT NP_BIND 183 190 GTP (BY SIMILARITY) .
Cc FT NP_BIND 265 269 GTP (BY SIMILARITY) .
Cc FT NP_BIND 323 326 GTP (BY SIMILARITY) .
Cc CONFLICT 76 76 E -> V (IN REF. 2).
Cc FT SEQUENCE 564 AA; 61232 MW; 423F7285FB9063E4 CRC64;
Cc
Query Match 4.9% Score 101; DB 1; Length 564;
Beet Local Similarity 26.1%; Fred. No. 1.2;
Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;

```

QY 114 KAAGHYVQVAFMKLAL--TGYVDHLTPLO--DMAAGRLDLAVAVEPVI FSDMEV 168  
 DB 211 --ADYVRPAIDLVILGEVGPVYTAGDVPAIDAGLKEAK-----NNVD 261  
 QY 169 KIITWGADTAACGDIISGLPVSARGSEIL----LGPADNFBGQKRLIAPTANSQOT 223  
 DB 262 VIM---DIAGRLIDKGMDELDKVKKFNLPTEVLVVDAMTQ--EAAAVTTFVVEI 315  
 QY 224 RGLIGCIITSLTGRDKNOVEGEVQVS 250  
 DB 316 -GITGALITLKDGDSDRGALSTVKEVS 341

## RESULT 13

DPO2\_MOUSE STANDARD; PRT; 600 AA.

AC P33611;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).  
 GN POLA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.  
 RX MEDLINE=93216788; PubMed=8463324;  
 RA Hayazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,  
 RA Hanaoka F.;

RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA  
 RT polymerase alpha-primase complex and their gene expression during  
 RT cell proliferation and the cell cycle."  
 RL J. Biol. Chem. 268:8111-8122(1993).  
 CC -1- FUNCTION: May play an essential role at the early stage of  
 CC chromosomal DNA replication by coupling the polymerase  
 CC alpha-primase complex to the cellular replication machinery (By  
 CC similarity).  
 CC -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme  
 CC (subunits A, B, C and D), which is assembled throughout the cell  
 CC cycle. The largest subunit (subunit A) has DNA polymerase  
 CC activity, the two smallest subunits (subunits C and D) have DNA  
 CC primase activity. Subunit B binds to subunit A.

CC -1- SUBCELLULAR LOCATION: Nucleus.  
 CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M  
 CC PHASE (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D13546; BAA02746.1; -  
 DR PIR: B46642; B46642.  
 DR MGI: MGI:99690; Polr2.  
 DR InterPro: IPR007200; DNA\_pol\_alpha\_B.  
 DR Pfam: PF04058; DNA\_pol\_alpha\_B\_1.  
 KW DNA replication; Nuclear protein; Phosphorylation.  
 FT DOMAIN 101  
 FT DOMAIN 107  
 FT DOMAIN 115  
 FT SEQUENCE 600 AA; 66267 MW; 79F94BEEF3FEBC CRC64;

Query Match 4.9%; Score 101; DB 1; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 1.3;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 105 LIRACMLVRKAAGHYVQVAFMKLALT-----CTYYVDHL-----TPQDMA 147

DB 27 LAELCVLRQTEDEGVSEILAFCTSGKTCUTVDILNFEYEVINKLKLKAMSHASKDSG 86  
 QY 148 HAGRLDLAAVEPV FSDMEVKIITWGADTAACGDI--ISGLP-----VSARGREI 197  
 DB 87 FAGTRDI-VSIQELLAEAEETLSTSTTSKGLKXVSTPEPLTKRVAARSPQ- 144  
 QY 198 LIGPADNFBGQKRLIAPTANSQOTRGLIACIITSLGRDKNOVEGEVQVSATQSGFL 257  
 DB 145 LISPSS-----FSPSATPSQK-----YTSRTNR-----GEVTTFGSAQ--- 178  
 QY 258 ATCVNAGVCTVPHAGSKTL--AGPKGITQKYNTPVDDLYG 297  
 DB 179 ----GLSMRGSGSVSLKVGDPPEPLTOSYKAMFOOLMG 215

## RESULT 14

Y447\_XYLFA STANDARD; PRT; 444 AA.

AC Q9PEF1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical zinc metalloprotease Xf1047 (EC 3.4.24.-).  
 GN Xf1047.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 NCBI\_TaxID=23711;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=945C;  
 RX MEDLINE=20165717; PubMed=10910347;  
 RA Simpson A.-J.G., Keimich F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carrer D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facincani A.P., Ferreira A.-J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Fullan L.R.,  
 RA Garner J.S., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohenstein J.D., Junqueira M.L., Kemper E.L., Kitejima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.P., Matukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montello-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.B.S.,  
 RA Nhami A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira P.G., Pereira H.A. Jr., Pasquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tuhako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";  
 RL Nature 406:151-159(2000).  
 CC -1- COFACTOR: Zinc (Probable).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (By similarity).  
 CC -1- SIMILARITY: Belongs to peptidase family M50B.  
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; AE003942; AAF3857.1; ALT_INIT.
DR PIR; G82728; G82728.
DR MEROPS; M50.004; -.
DR InterPro; IPR001476; PDZ.
DR InterPro; IPR004387; Pept_M50_Zn.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR008915; Peptidase_M50.
DR Pfam; PF02163; Peptidase_M50; 1.
DR SMART; SM00228; PDZ; 2.
DR TIGRfams; TIGR00054; TIGR00054; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
  Inner membrane; Complete proteome.
FT METAL 22 22 ZINC (CATALYTIC) (POTENTIAL).
FT ACT SITE 23 23 ZINC (CATALYTIC) (POTENTIAL).
FT METAL 26 26 ZINC (CATALYTIC) (POTENTIAL).
FT TRANSMEM 98 120 POTENTIAL.
FT TRANSMEM 371 393 POTENTIAL.
FT TRANSMEM 418 440 POTENTIAL.
FT DOMAIN 192 276 PDZ.
SQ SEQUENCE 444 AA; 48410 MW; B10D1532DBA4A34 CRC64;

Query Match 4.6%; Score 94; DB 1; Length 444;
Best Local Similarity 20.4%; Pred. No. 3.3;
Matches 64; Conservative 52; Mismatches 117; Indels 80; Gaps 16;

8 AVFTGLALLTSPYKVLALILMGLITREVAH;QWVLPNVEGRDAIILLCAV 67
: : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : :
93 SWQRIALVAGPLANILLMLLVLEVIKQVSAIV-----GRAEHLAAGI 143
: : : : : : : : : : : : : : : : : : : : : : : : : :
68 HPELIPDITKLLAIIFGLMVLQAGITKVFYVPAQGLIRACMLVRAAG---GHVQM 123
: : : : : : : : : : : : : : : : : : : : : : : : : :
144 HP-----GDRITAIIDG-----RQVTSWSASMLITAAANDRONAVLRVIGPYGER 188
: : : : : : : : : : : : : : : : : : : : : : : : : :
124 A--EMKLAALGTGYVDHLTPLO--DNVAHAGRLAVAVEVIVSDMEVK-----IITWG 174
: : : : : : : : : : : : : : : : : : : : : : : : : :
189 SEHTLESKILKQPFDERHVALVGINWQMLQPIIAKIEBGSIAEGAIKFGDIVLAVDG 248
: : : : : : : : : : : : : : : : : : : : : : : : : :
175 ADTACAGDIIS-----GLP--VSARGREIL--LGPADNPEGQWRL---LAPI 216
: : : : : : : : : : : : : : : : : : : : : : : : : :
249 QQTISTEDLVNQIKLGRDHGPMIEIRRGERRALALSPKSAQGV-WLGVKTNQPV 307
: : : : : : : : : : : : : : : : : : : : : : : : : :
217 TAY-SQGRGLIGCITSL--TGR-----DKQVGEVQVYSTATQSF 256
: : : : : : : : : : : : : : : : : : : : : : : : : :
308 PAFDSQQRVGYLAIVPLAIRETGWTADSLGMMKRRIITGQASAKNISGISIAKIAN--- 364
: : : : : : : : : : : : : : : : : : : : : : : : : :
257 LATCVNGVCWTFV 269
: : : : : : : : : : : : : : : : : : : : : : : : : :
365 -ASAKRGVCWTFV 376
: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
CYAA LEIDO STANDARD; PRT; 1380 AA.
ID CYAA LEIDO STANDARD; PRT; 1380 AA.
AC Q27675;
DT 15-JUL-1998 (Rel. 35, Last Created)
DT 15-JUL-1998 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Receptor-type adenylylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-
  lyase) (Adenylyl cyclase).
GN RAC-A.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=IS Sudanese;
RX MEDLINE=95340554; PubMed=7615561;
RA Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.;
RT "A family of putative receptor-adenylylate cyclases from Leishmania
  donovani."

```

```

RL J. Biol. Chem. 270:17551-17558(1995).
CC -1- FUNCTION: Could act as a receptor for a unknown ligand.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote)
  but not in the mammalian host stage of the parasite life cycle.
CC -1- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U17042; AAA74998.1; -.
DR PIR; T18309; T18309.
DR InterPro; IPR001054; G_cyclase.
DR Pfam; PF00211; guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS0125; GUANYLATE CYCLASES 2; 1.
DR Lyase; CAMP biosynthesis; Transmembrane; Receptor; Glycoprotein;
  Metal-binding; Magnesium.
FT METAL 1 34 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 35 55 POTENTIAL.
FT DOMAIN 56 891 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 892 912 POTENTIAL.
FT DOMAIN 913 1380 CYTOPLASMIC (POTENTIAL).
FT METAL 918 938 MAGNESIUM (BY SIMILARITY).
FT METAL 981 981 MAGNESIUM (BY SIMILARITY).
FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1380 AA; 151692 MW; 6B2D5F73C1107A0 CRC64;

Query Match 4.6%; Score 93.5; DB 1; Length 1380;
Best Local Similarity 21.0%; Pred. No. 15;
Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;

55 GGRDAIILLTCAVHP---ELIFDITKL-----LTAIFGLM-----V 88
: : : : : : : : : : : : : : : : : : : : : : : : : :
101 GGRPIKIL-----HPPDDQNLVDAIEVLHSLARGKELAVLGYLDGRILTAALSNADV 155
: : : : : : : : : : : : : : : : : : : : : : : : : :
89 LQNGITKVP-----YVPAQGLIRACMLVRAAGHYV-----QMAENK 127
: : : : : : : : : : : : : : : : : : : : : : : : : :
156 VQSGMLVLAIFFTSGSGVRTWSDSVYFTRAEPVLEKVLVLM---HIVRLARRVAFWR 210
: : : : : : : : : : : : : : : : : : : : : : : : : :
128 LAALTG-----TYVDHLTPLOMAHAGRLAVA--VEPVIFSDMEVK----- 169
: : : : : : : : : : : : : : : : : : : : : : : : : :
211 ---LTGMEHFGESLITVQDTLTSL-----LRDPAVLXTVYSESVAVDEEAPNAMD 260
: : : : : : : : : : : : : : : : : : : : : : : : : :
170 -----IITWGAADTACAGDIISGLPVARGREILIG-----PADNFEQ 208
: : : : : : : : : : : : : : : : : : : : : : : : : :
261 TNPQVITVMAAPQVITYLEKVLTPRTSSAYVISCMIQVPEVRYRLISAGSIKQ 320
: : : : : : : : : : : : : : : : : : : : : : : : : :
209 GWRLLAPITNVSQGTGGLGICITSLTGRDKQVGE-----QVYVSTATQSFATQVNV 264
: : : : : : : : : : : : : : : : : : : : : : : : : :
321 DGRILASATV-----SVSGEGIKTMEVLAQNSVIE---NSG 356
: : : : : : : : : : : : : : : : : : : : : : : : : :
265 CWTVFAGAGSKTLAQRK---GPITQMTNVVDQDLVGWQAPPARSMTPTCGSSDLYLV 320
: : : : : : : : : : : : : : : : : : : : : : : : : :
357 SPFYVDDSTERTLGRKASEAPLSRKYT-VDE---FQAHPSIAKLMALGWLSTLVQO 412
: : : : : : : : : : : : : : : : : : : : : : : : : :
321 TRHADVIIPRRRDSGSLSRPV---SYLNGSGGPILC--PSGHAVGIFFAAVCTRGV 376
: : : : : : : : : : : : : : : : : : : : : : : : : :
413 TLEQTDIVNRSTYKAGLFNRQNRFPVIGDVIADYGGP--CEFLAQLFG--ASCYCNQGG 468
: : : : : : : : : : : : : : : : : : : : : : : : : :
377 AKAV-----DFIVESME 389
: : : : : : : : : : : : : : : : : : : : : : : : : :
469 HSSILTVLQNASWDIVPDSFK 490
: : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 6, 2004, 09:31:50  
Job time : 9.20459 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 36.7566 Seconds  
(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MASCGAGATGATLTLSP.....RGVAKVDPIPVSEYETTR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.25.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_procent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_archaeap:\*  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1986	96.7	3010	12 Q9J3F9	Q9J3F9 hepatitis c
2	1981	96.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
3	1979	96.4	3010	12 Q9DTR6	Q9DTR6 hepatitis c
4	1974	96.2	3010	12 Q9DTR6	Q9DTR6 hepatitis c
5	1973	96.1	3010	12 Q9DTR6	Q9DTR6 hepatitis c
6	1973	96.1	3010	12 Q9DTR6	Q9DTR6 hepatitis c
7	1971	96.0	3010	12 Q9J3H3	Q9J3H3 hepatitis c
8	1970	96.0	3010	12 Q9J3H3	Q9J3H3 hepatitis c
9	1968	95.9	1166	12 Q81755	Q81755 hepatitis c
10	1968	95.9	12284	12 Q81817	Q81817 hepatitis c
11	1968	95.9	3010	12 P89966	P89966 hepatitis c
12	1968	95.9	3010	12 Q9J3G6	Q9J3G6 hepatitis c
13	1967	95.8	3010	12 Q9J3A2	Q9J3A2 hepatitis c
14	1967	95.8	3010	12 Q9J3D7	Q9J3D7 hepatitis c
15	1967	95.8	3010	12 Q9Q1X6	Q9Q1X6 hepatitis c
16	1967	95.8	3010	12 Q9Q1X5	Q9Q1X5 hepatitis c

17	1966	95.8	3008	12 Q9J3F4	Q9J3F4 hepatitis c
18	1965	95.7	3010	12 Q9J3H0	Q9J3H0 hepatitis c
19	1964	95.7	3010	12 Q9J3H9	Q9J3H9 hepatitis c
20	1963	95.6	3010	12 Q9J3H9	Q9J3H9 hepatitis c
21	1963	95.6	3010	12 Q81760	Q81760 hepatitis c
22	1961	95.5	3010	12 Q9Q1Y3	Q9Q1Y3 hepatitis c
23	1961	95.5	3014	12 Q9DTR0	Q9DTR0 hepatitis c
24	1960	95.5	3010	12 P88803	P88803 hepatitis c
25	1960	95.5	3010	12 Q9J3H0	Q9J3H0 hepatitis c
26	1960	95.5	3010	12 Q9J3H5	Q9J3H5 hepatitis c
27	1959	95.4	3010	12 Q9Q1X8	Q9Q1X8 hepatitis c
28	1959	95.4	3010	12 Q9Q1X7	Q9Q1X7 hepatitis c
29	1958	95.4	3010	12 Q9Q1Y5	Q9Q1Y5 hepatitis c
30	1958	95.4	3010	12 Q9DTR9	Q9DTR9 hepatitis c
31	1958	95.4	3010	12 Q9J3H6	Q9J3H6 hepatitis c
32	1957	95.3	3010	12 Q9DTR0	Q9DTR0 hepatitis c
33	1955	95.2	3010	12 Q68826	Q68826 hepatitis c
34	1954	95.2	3015	12 Q9WPH5	Q9WPH5 hepatitis c
35	1952	95.1	3010	12 Q9Q1Y4	Q9Q1Y4 hepatitis c
36	1951	95.0	3010	12 Q9J3I1	Q9J3I1 hepatitis c
37	1951	95.0	3010	12 Q9WXX2	Q9WXX2 hepatitis c
38	1951	95.0	3011	12 Q9DTR3	Q9DTR3 hepatitis c
39	1950	95.0	3010	12 Q09796	Q09796 hepatitis c
40	1949	94.9	3010	12 Q9DTR5	Q9DTR5 hepatitis c
41	1947	94.8	3010	12 Q9WIK8	Q9WIK8 hepatitis c
42	1947	94.8	3010	12 Q9Q1Y6	Q9Q1Y6 hepatitis c
43	1946	94.8	1275	12 Q06642	Q06642 hepatitis c
44	1946	94.8	3010	12 Q9Q1Y7	Q9Q1Y7 hepatitis c
45	1946	94.8	3010	12 Q9Q1Y8	Q9Q1Y8 hepatitis c

## ALIGNMENTS

RESULT 1  
ID Q9J3F9 PRELIMINARY; PRT; 3010 AA.

AC Q9J3F9;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_Taxid=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MD33;  
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease progression."  
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL AF207774; AAF65964.1; --  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.  
DR PIR; P50329; P50329.  
DR HSSP; P27958; 1HE1.  
DR MEROS; S29.001; --  
DR MEROS; U39.001; --  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005824; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003768; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.



DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR003045; Cys\_Ser\_trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_core.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PsVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR Pfam: PF018602; HCV\_NS1; 1.  
 DR SMART: SMO0487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SO SEQUENCE 3010 AA, 327102 MW, 716209DB33E60C CRC64.

Query Match 96.7%; Score 1986; DB 12; Length 3010;  
 Best Local Similarity 94.9%; Pred. No. 5,4e-157;  
 Matches 373; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

DR 1 MAASCGAVVIGLALLTSPYKVLARLIWLOYLITVTEALQVWIPPLVNRGRDAI 60  
 DB 814 MAASCGAVVIGLALLTSPYKVLARLIWLOYLITVTEALQVWIPPLVNRGRDAI 873  
 QY 61 ILTCAVHPELIDITKLLAIFGRLVQAGITKVPYVRAQGLIRACMLVRKAGHY 120  
 DB 874 ILTCAVHPELIDITKLLAIFGRLVQAGITRMPYFRAQGLIRACMLVRKAGHY 933  
 QY 121 VQMAFKKLALNGTYVYDHLTPQDMAHAGRLAVAVEVIFSDMEVKIITWGADTAC 180  
 DB 934 VQMAFKKLALNGTYVYDHLTPLRDWAHAGRLAVAVEVIFSDMETKIITWGADTAC 993  
 QY 181 GDIISGLPVARSRRREILLGPADNPEGQWRLAPITAAVSOQRLGLGCIITSLNRDN 240  
 DB 994 GDIISGLPVARSRRREILLGPADNPEGQWRLAPITAAVSOQRLGLGCIITSLNRDN 1053  
 QY 241 QVEGEVQVSTAFQSLATVNGVCVTFVHAGSKTLAKGPIITMYNNVQDILWGQA 300  
 DB 1054 QVEGEVQVSTAFQSLATVNGVCVTFVHAGSKTLAKGPIITMYNNVQDILWGQA 1113  
 QY 301 PGARSMTPCTGSSPLVLTTRADVIYVRRRDSGSLSPPVSYLGGSSGGLPLCS 360  
 DB 1114 PGARSMTPCTGSSPLVLTTRADVIYVRRRDSGSLSPPVSYLGGSSGGLPLCS 1173  
 QY 361 GHAAGIFRAAVCTRGVAKAVDPIVDSMETTNR 393  
 DB 1174 GHAAGIFRAAVCTRGVAKAVDPIVDSMETTNR 1206

RESULT 2  
 ID 09J3H7 PRELIMINARY; PRT, 3010 AA.  
 AC 09J3H7;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 DE Hepatitis C virus..  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepatitis C virus.  
 CC NCBI\_Taxid=1103;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD5;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression".  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AF207756; AAF65946.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: P00804; P00804.  
 DR PIR: P03329; P03329.  
 DR HSP: P26663; IUXP.  
 DR GO:0016021; C:Integral to membrane; IEA.  
 DR GO:0019028; C:Viral capsid; IEA.  
 DR GO:0019031; C:Viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:Electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR000345; Cys\_Ser\_trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PsVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR Pfam: PF018602; HCV\_NS1; 1.  
 DR SMART: SMO0487; DEXDC; 1.

QY	1	MAASCGGAVFGLALITSPYKYKLALRIWLOYLITRVZAHQWMTPLNVRGGRDAI	60
Db	814	MAASCGGAVFGLALITSPYKYFLARLIWLOYLITRVZAHQWMTPLNVRGGRDAI	873
QY	61	ILLCAVPELIFDITKLALIFGPIWLOAGITVPYFVFAOGILRACMLVRKAAGHY	120
Db	874	ILLCAVHPSGLIFDITKLALIFGPIWLOAGITRMPFVFAOGILRACMLVRKAAGHY	933
QY	121	VQAFPMKLAALTGYVYDHLTPELOMAHAGLRDAVAVEPIFSDMEXKIIITWQADTAAC	180
Db	934	IQAFPMKLAALTGYVYDHLTPELOMAHAGLRDAVAVEPIFSDMEXKIIITWQADTAAC	993
QY	181	GDITSGIPVSRARREGLLGGADNPEGGMPLAIPITVSOOTGSLGCIITSLTGRDX	240
Db	994	GDITGLPVSARRREGLLGGADNPEGGMPLAIPITVSOOTGSLGCIITSLTGRDX	1053
QY	241	QVEGEVQVSTATOSFLATCVGVCMVTFHGAGSKTLAGPKGPITOMYTNVDOLVQQA	300
Db	1054	QVEGEVQVSTATOSFLATCVGVCMVTFHGAGSKTLAGPKGPITOMYTNVDOLVQQA	1113
QY	301	PPGARSMTPCTCGSSDLVLTNRADVITVRRRGDSRGLSLSPRPSTYKSGSSGPELCP	360
Db	1114	PPGARSMTPCTCGSSDLVLTNRADVITVRRRGDSRGLSLSPRPSTYKSGSSGPELCP	1173
QY	361	GHAAGIFPAAVCTRGVAKAVFIPESMETMR	393
Db	1174	GHAAGIFPAAVCTRGVAKAVFIPESMETMR	1206

Query Match 96.5%, Score 1981, DB 12, Length 3010;  
 Best Local Similarity 94.7%, Pred. No. 1,4e-156;  
 Matches 372, Conservative 11, Mismatches 10, Indels 0, Gaps 0;

PROSITE:PS00190; CYTOCHROME C; 1.  
 KM Cost protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 RM; D8653F7317FFA106 CR644;

EMBL; AB049091, BAB18804.1; -.  
 PIR; A61196, A61196.  
 PIR; PS0329, PS0329.  
 HSSP; P26653, 1UXP.  
 GO; GO:0016021; C: integral to membrane; IEA.  
 GO; GO:0019028; C: viral capsid; IEA.  
 GO; GO:0019031; C: viral envelope; IEA.  
 GO; GO:0005524; F: ATP binding; IEA.  
 GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 GO; GO:0005489; F: electron transporter activity; IEA.

RESULT 3  
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 Q9DTE6  
 ID 01-MAR-2001 (TREMBlrel. 16, Created)  
 AC 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 DE Hepatitis C virus.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CX NCBI\_Taxid=11103;  
 CY (1)  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=HCV1142;  
 RC Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 with hepatocellular carcinoma: the 'progression score' revisited.";  
 RL Submitted (Sep-2000) to the EMBL/Genbank/DDBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MNNA (BY SIMILARITY).

	DR	GO:	GO:0016787;	F:hydrolase activity; IEA.
	DR	GO:	GO:0003723;	F:RNA binding; IEA.
	DR	GO:	GO:0003968;	F:RNA-directed RNA polymerase activity; IEA.
	DR	GO:	GO:0008326;	F:serine-type peptidase activity; IEA.
	DR	GO:	GO:0005198;	F:structural molecule activity; IEA.
	DR	GO:	GO:0016740;	F:transferase activity; IEA.
	DR	GO:	GO:0006118;	P:electron transport; IEA.
	DR	GO:	GO:0006508;	P:proteolysis and peptidolysis; IEA.
	DR	GO:	GO:0006350;	P:transcription; IEA.
	DR	GO:	GO:0019079;	P:viral genome replication; IEA.
	DR	GO:	GO:0019087;	P:viral transformation; IEA.
	DR	InterPro:	IPR009003;	Cys_ser_lypsin.
	DR	InterPro:	IPR000345;	Cytic_heme_BS.
	DR	InterPro:	IPR001410;	DEAD.
	DR	InterPro:	IPR002522;	HCV capsid.
	DR	InterPro:	IPR002521;	HCV env.
	DR	InterPro:	IPR002519;	HCV NS1.
	DR	InterPro:	IPR002511;	HCV NS2.
	DR	InterPro:	IPR000745;	HCV NS4a.
	DR	InterPro:	IPR001490;	HCV NS4b.
	DR	InterPro:	IPR002868;	HCV NS5a.
	DR	InterPro:	IPR002166;	HCV RdRp.
	DR	InterPro:	IPR001650;	Helicase_C.
	DR	InterPro:	IPR004109;	peptidase_C29.
	DR	InterPro:	IPR007095;	RNA_pol_DS_PS.
	DR	InterPro:	IPR007094;	RNA_pol_PSVLr.
	DR	Pfam:	PF01543;	HCV capsid; 1.
	DR	Pfam:	PF01542;	HCV env; 1.
	DR	Pfam:	PF01539;	HCV env; 1.
	DR	Pfam:	PF01560;	HCV NS1; 1.
	DR	Pfam:	PF01538;	HCV NS2; 1.
	DR	Pfam:	PF02907;	HCV NS3; 1.
	DR	Pfam:	PF01006;	HCV NS4a; 1.
	DR	Pfam:	PF01001;	HCV NS4b; 1.
	DR	Pfam:	PF01506;	HCV NS5a; 1.
	DR	Pfam:	PF002271;	helicase_C; 1.
	DR	Pfam:	PF00998;	viral_RdRp; 1.
	DR	PRODOM:	PD186062;	HCV_NS1; 1.
	DR	SMART:	SMO0487;	DExDC; 1.
	DR	SMART:	SMO0490;	HELICC; 1.
	DR	PROSITE:	PS00190;	CYCLOCHROME_C; 1.
	KW	Atp-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;		
	KM	Hydrolase; Nonstructural protein; Polypeptide; Transmembrane.		
	SO	SEQUENCE	3010 AA; 327042 MW; 3807DC6879684C95 CRC64;	
		Query Match	96.4%; Score 19799; DB 12; Length 3010;	
		Best Local Similarity	94.4%; Pred. No. 2.1e-156;	
		Matches 371; Conservative 11; Mismatches 11; Indels	0; Gaps	0;
OY		1	MASAGCAVFGCLALTLSPYKYLRLIWLQYLITRRZAHQWIMPIPLNVGGSDAI	60
DB		814	MAACGGGVFGLVLTLSPPYKVFIAKLIMWIQYITFEAAHLQVWVPPLNVGGRDAI	873
OY		61	IILCAVHELIIFITKLILAIFGPLVLWLAGITRVFYFAAQGLIRACMLVRRAAGHY	120
DB		874	IILICAHELIIFITKLILAIFGPLVLWLAGITRVFYFAAQGLIRACMLVRRAAGHY	933
OY		121	VQMAFMKLAALTGYVVDHLLPLODMAHAIGLDIAVAVEPYISDMEYKITTWGADTAA	180
DB		934	VQMAFMKLAALTGYVVDHLLPLEMDMAHTGIRDIAVAVEPVFDMEIKITTWGADTAA	993
OY		181	GDIISGLPVASRRREITLGPANDPEOGMWLAPITAVSQOTSGLGCIITSITGRDKN	240
DB		994	GDIISGLPVASRRREITLGPANDPEOGMRLLAPITAVSQOTSGLGCIITSITGRDKN	1053
OY		241	QVEGEVOVVSSTATSFLATCVNGVCWTVPFHAGSKTLAGPKGITTYNTVNDDLVGQA	300
DB		1054	QVEGEVOVVSSTATSFLATCVNGVCWTVPFHAGSKTLAGPKGITTYNTVNDDLVGQA	1113
OY		301	PRGARSMTPCCSSDLIVYTRHADVIPIRRRGSRSLSPRVSYLKGSSGGPELLCP	360

Db 1114 PGARSLPCTCGSSDLYLVRHADVIPVRRGDSRLSPRVSYLKSGGFLPCPS 1173

QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393

Db 1174 GHVGVIFRAAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 4

ID 09DTE6 PRELIMINARY; PRT; 3010 AA.

AC 09DTE6; 01-MAR-2001 (TRENBLrel. 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Genome polypeptidein.

OS Hepatitis C virus.

OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

CC Hepacivirus.

CC NCBI\_taxid=11103;

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN=HCV1221;

CC Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,

CC Harabara T., Ohba Y., Kanai K., Maruo H., Baba K., Hijioka M.,

CC Mishihiro S.,

CC "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients

CC with hepatocellular carcinoma: the 'progression score' revisited.",

CC Submitted (SPP-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA (BY SIMILARITY).

CC EMBL; AB049101; BAB18814.1; -.

CC PIR; A61196; A61196.

CC PIR; PQ0246; PQ0246.

CC PIR; PS0329; PS0329.

CC HSSP; P26663; IJXP.

CC GO; GO:0016021; C:integral to membrane; IEA.

CC GO; GO:0019028; C:viral capsid; IEA.

CC GO; GO:0019031; C:viral envelope; IEA.

CC GO; GO:0005524; F:ATP binding; IEA.

CC GO; GO:0008026; F:ATP dependent helicase activity; IEA.

CC GO; GO:0005489; F:electron transporter activity; IEA.

CC GO; GO:0016787; F:hydrolase activity; IEA.

CC GO; GO:0003723; F:RNA binding; IEA.

CC GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

CC GO; GO:0008236; F:serine-type peptidase activity; IEA.

CC GO; GO:0005198; F:structural molecule activity; IEA.

CC GO; GO:0016740; F:transferase activity; IEA.

CC GO; GO:0006508; F:electron transport; IEA.

CC GO; GO:0006350; F:proteolysis and peptidolysis; IEA.

CC GO; GO:0019079; F:viral genome replication; IEA.

CC GO; GO:0019087; F:viral transformation; IEA.

CC Interpro: IPR009003; Cys\_Ser\_Typsin.

CC Interpro: IPR003045; CytC\_heme\_BS.

CC Interpro: IPR001410; DEAD.

CC Interpro: IPR002523; HCV\_capsid.

CC Interpro: IPR002521; HCV\_core.

CC Interpro: IPR002519; HCV\_env.

CC Interpro: IPR002531; HCV\_NS1.

CC Interpro: IPR002518; HCV\_NS2.

CC Interpro: IPR000745; HCV\_NS4a.

CC Interpro: IPR001490; HCV\_NS4b.

CC Interpro: IPR002868; HCV\_NS5a.

CC Interpro: IPR002166; HCV\_NS5b.

CC Interpro: IPR001650; Helicase\_C.

CC Interpro: IPR004109; Peptidase\_C29.

CC Interpro: IPR007095; RNA\_pol\_DS\_PS.

CC Interpro: IPR007094; RNA\_pol\_PSVir.

CC Pfam; PF01543; HCV\_capsid.1.

CC Pfam; PF01542; HCV\_core.1.

CC Pfam; PF01539; HCV\_env.1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR ProDom; PD186062; HCV\_NS1; 1.

DR SMART; SM00487; DEXDC1; 1.

DR PROSITE; PS00190; CYTOCHROME\_C; 1.

KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;

KW Hydroxylase; Nonstructural protein; Polypeptidein;

KW RNA-directed RNA polymerase; Transferase; Transmembrane.

SQ SEQUENCE 3010 AA; 327108 MW; DE182D810BF78EE4 CRC64;

Query Match 96.2% Score 1974; DB 12; Length 3010;

Best Local Similarity 94.9% Pred. No. 5.4e-156;

Matches 373; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKYLARLIMLYITRVAHLQVIPPINAGGRDAI 60

Db 814 MAASCGAVFGLVLTLSPPYKVFARLIMLYITRVAHLQVIPPINAGGRDAI 873

QY 61 ILTCAVHPELIPDTIKLLAIPGPIWLTQGIKRVYFVPAQGLISACMLVRAAGHY 120

Db 874 ILTCAVHPELIPDTIKLLAIPGPIWLTQGIKRVYFVPAQGLISACMLVRAAGHY 933

QY 121 VQMAFMKLAALTGIVYDHLTPLODMAHAGRLDAVAPEYIFSDMEVKITWGADTAAC 180

Db 934 VQMAFMKLAALTGIVYDHLTPLODMAHAGRLDAVAPEYIFSDMEVKITWGADTAAC 993

QY 181 GDIIISGLFVSARRREITLPGADNPEGGRLLPITAYISQTRGLIGCITLSITGDKN 240

Db 994 GDIIISGLFVSARRREITLPGADNPEGGRLLPITAYISQTRGLIGCITLSITGDKN 1053

QY 241 QVEGEVQVSTATOSFLATCVNCGVTFHAGASKTLAGPKPITQWTVNDQDLVMOA 300

Db 1054 QVEGEVQVSTATOSFLATCVNCGVTFHAGASKTLAGPKPITQWTVNDQDLVMOA 1113

QY 301 PGARSLPCTCGSSDLYLVRHADVIPVRRGDSRLSPRVSYLKSGGFLPCPS 360

Db 1114 PGARSLPCTCGSSDLYLVRHADVIPVRRGDSRLSPRVSYLKSGGFLPCPS 1173

QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393

Db 1174 GHVGVIFRAAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 5

ID 09DTE4 PRELIMINARY; PRT; 3010 AA.

AC 09DTE4; 01-MAR-2001 (TRENBLrel. 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Genome polypeptidein.

OS Hepatitis C virus.

OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

CC Hepacivirus.

CC NCBI\_taxid=11103;

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN=HCV1221;

CC Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,

CC Harabara T., Ohba Y., Kanai K., Maruo H., Baba K., Hijioka M.,

CC Mishihiro S.,

CC "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients

CC with hepatocellular carcinoma: the 'progression score' revisited.",

CC Submitted (SPP-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL; AB049093; BAB1806.1; -.

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; P00804; P00804.

DR PIR; P0329; P0329.

DR HSSP; P26663; IUXP.

DR GO; GO:0016021; C: integral to membrane; IEA.

DR GO; GO:0019028; C: viral capsid; IEA.

DR GO; GO:0019031; C: viral envelope; IEA.

DR GO; GO:0005524; F: ATP binding; IEA.

DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F: electron transporter activity; IEA.

DR GO; GO:0003723; F: RNA binding; IEA.

DR GO; GO:0008236; F: RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0005198; F: structural molecule activity; IEA.

DR GO; GO:0016740; F: transferase activity; IEA.

DR GO; GO:0006118; P: electron transport; IEA.

DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P: transcription; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.

DR GO; GO:0019087; P: viral transformation; IEA.

DR InterPro; IPR003003; Cys Ser trypsin.

DR InterPro; IPR000345; CytC\_heme\_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR000745; HCV\_NS4.

DR InterPro; IPR001490; HCV\_NS4B.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_NS5a.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4; 1.

DR Pfam; PF01001; HCV\_NS4B; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RDRP; 1.

DR ProDom; PD186062; HCV\_NS1; 1.

DR SMART; SM00487; DEXDC1; 1.

DR PROSITE; PS00190; CYTOCHROME\_C; 1.

KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.

QO POLYPROTEIN; 3010 AA; 327324 MW; 3DBCF249BD1151C CR64;

Query Match 96.1%; Score 1973; DB 12; Length 3010;

Best Local Similarity 94.4%; Pred. No. 6.6e-156;

Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 1 MAASCGAVFTGLALLTSPYKYLARLIMWLOYLITRVAHLQWIPPLNVRGGDAI 60

DB 814 MAASCGAVFTGLALLTSPYKYLARLIMWLOYLITRVAHLQWIPPLNVRGGDAI 873

QY 61 ILTCAVHEPILFITKLLAIFGLNWLQAGIKVYFPAQGLPAQCLVRAAGGHY 120

DB 874 ILTCAVHEPILFITKLLAIFGLNWLQAGIKVYFPAQGLPAQCLVRAAGGHY 933

QY 121 VQNAFMKLAALGTYYVDHLLTPIQDMAHAGIRLAVAVEPIFSDMEVKITWGADTAAC 180

DB 934 VQNAFMKLAALGTYYVDHLLTPIQDMAHAGIRLAVAVEPIFSDMEVKITWGADTAAC 993

QY 181 GDIIGLPSVARGREILLGPADNFEQGWRLAPITAYSOOTRGLCCITSLTGRPN 240

DB 994 GDIIGLPSVARGREILLGPADNFEQGWRLAPITAYSOOTRGLCCITSLTGRPN 1053

QY 241 QVEGEVQVVSATQSFATCVNGVCMTFVHGAGSKTLGAPKPIPTOMYTNNDQIVGQA 300

DB 1054 QVEGEVQVVSATQSFATCVNGVCMTFVHGAGSKTLGAPKPIPTOMYTNNDQIVGQA 1113

QY 301 PGASMSPTCTGSSDLYLVRHADVIVRRRDSRGSLSPRVSYLKSGSGGSLCPG 360

DB 1114 PGASMSPTCTGSSDLYLVRHADVIVRRRDSRGSLSPRVSYLKSGSGGSLCPG 1173

QY 361 GHAVGIFRAVCTRGVAVKADPIPVESMETMR 393

DB 1174 GHAVGIFRAVCTRGVAVKADPIPVESMETMR 1206

RESULT 6

0807P3 PRELIMINARY; PRT; 3010 AA.

AC 0807P3

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Polypeptide.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepatitis.

ON NCBI\_Taxid=11103;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MLE.

RX MEDLINE-22047193; PubMed-12051758;

RA Kishine H., Sugiyama K., Hijioka M., Kato N., Takahashi H., Nochi T., Nio Y., Hosaka M., Miyamari Y., Shimotohno K.,

RT "Subgenomic replicon derived from a cell line infected with the hepatitis C virus."

RL Biochem. Biophys. Res. Commun. 293:993-999 (2002).

DR EMBL; AB080299; BAC54896.1; -.

DR GO; GO:0019028; C: viral capsid; IEA.

DR GO; GO:0019031; C: viral envelope; IEA.

DR GO; GO:0005524; F: ATP binding; IEA.

DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F: electron transporter activity; IEA.

DR GO; GO:0003723; F: RNA binding; IEA.

DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F: structural molecule activity; IEA.

DR GO; GO:0005198; F: structural molecule activity; IEA.

DR GO; GO:0006118; P: electron transport; IEA.

DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P: transcription; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.

DR GO; GO:0019087; P: viral transformation; IEA.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR000345; CytC\_heme\_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR000745; HCV\_NS4.

DR InterPro; IPR001490; HCV\_NS4B.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_NS5a.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR Pfam: PF01506; HCV NS1; 1.  
 DR SMART: SMO0487; DEXDC; 1.  
 DR SMART: SMO0490; Helicase C; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR Polyprotein.  
 SQ SEQUENCE 3010 AA; 327097 MW; EE6418CA723E686 CRC64;

Query Match 96.1%; Score 1973; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 6.6e-156;  
 Matches 372; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIMWLOYLITRVAHLQVWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFGLALTLSPYKVLARLIMWLOYLITRVAHLQVWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVHPELLIDITKLLAIFGLPLMTAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 874 ILITCAVHPELLIDITKLLAIFGLPLMTAGITKVPYFRAQGLIRACMLVRKAGGHY 933  
 QY 121 VQMAFMKLAALTGYVVDHLPLODMNAHGRDLAVAVEVPSDMVKIITWGADTAAC 180  
 DB 934 VQMAFMKLAALTGYVVDHLPLODMNAHGRDLAVAVEVPSDMVKIITWGADTAAC 993  
 QY 181 GDIIISGLPVSARRREIILGPADNFEQGRRLIAPITAVSQOTRGLIGCIITSLGRDN 240  
 DB 994 GDIIISGLPVSARRREIILGPADNFEQGRRLIAPITAVSQOTRGLIGCIITSLGRDN 1053  
 QY 241 QVGEVQVSTATOSFATCNGVCTVFEHAGSKTLAGKPIITOMYNNVDDLVGMOA 300  
 DB 1054 QVGEVQVSTATOSFATCNGVCTVFEHAGSKTLAGKPIITOMYNNVDDLVGMOA 1113  
 QY 301 PPGARSMTPCTCGSSDLYLTRADYIPVRRGDSRGLSPRPVSYLKGSGGFLCPSS 360  
 DB 1114 PPGARSMTPCTCGSSDLYLTRADYIPVRRGDSRGLSPRPVSYLKGSGGFLCPSS 1173  
 QY 361 GHAVGIFRAVCTRGVAKANDPFPVSMETMR 393  
 DB 1174 GHAVGIFRAVCTRGVAKANDPFPVSMETMR 1206

## RESULT 7

Q9J3H3 PRELIMINARY; PRT; 3010 AA.

AC Q9J3H3; PRELIMINARY; PRT; 3010 AA.  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=1103;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MD19;  
 RA Nageyama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: AF207760; AA65590.1; -;  
 DR PIR: A61196; A61196.  
 DR PIR: PS0329; PS0329.

DR HSP; P26663; IUXP.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transporter activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: structural molecule activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006118; F: electron transport; IEA.  
 DR GO: GO:0006350; F: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0019079; F: viral genome replication; IEA.  
 DR GO: GO:0019087; F: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser LysPn.  
 DR InterPro: IPR003445; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR00745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR Pfam: PF01506; HCV NS1; 1.  
 DR ProDom: PD186062; DEXDC; 1.  
 DR SMART: SMO0487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327234 MW; 44C3467649C88D CRC64;

Query Match 96.0%; Score 1971; DB 12; Length 3010;  
 Best Local Similarity 93.9%; Pred. No. 9.7e-156;  
 Matches 369; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIMWLOYLITRVAHLQVWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFGLALTLSPYKVLARLIMWLOYLITRVAHLQVWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVHPELLIDITKLLAIFGLPLMTAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 874 ILITCAVHPELLIDITKLLAIFGLPLMTAGITKVPYFRAQGLIRACMLVRKAGGHY 933  
 QY 121 VQMAFMKLAALTGYVVDHLPLODMNAHGRDLAVAVEVPSDMVKIITWGADTAAC 180  
 DB 934 VQMAFMKLAALTGYVVDHLPLODMNAHGRDLAVAVEVPSDMVKIITWGADTAAC 993  
 QY 181 GDIIISGLPVSARRREIILGPADNFEQGRRLIAPITAVSQOTRGLIGCIITSLGRDN 240  
 DB 994 GDIIISGLPVSARRREIILGPADNFEQGRRLIAPITAVSQOTRGLIGCIITSLGRDN 1053  
 QY 241 QVGEVQVSTATOSFATCNGVCTVFEHAGSKTLAGKPIITOMYNNVDDLVGMOA 300

DB 1054 QVEGEVQVSTAFATGCVTGVHGAQAKTLAGKPIITQMTNTVDODLVGMOS 1113  
 QY 301 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCP 360  
 DB 1114 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCP 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 8  
 Q68788  
 ID Q68788 PRELIMINARY; PRT; 3010 AA.  
 AC Q68788;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
 DE HCV polyprotein (Genome polyprotein).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9636218; PubMed=8720135;  
 RA Seki M., Honda Y.,  
 RT "Phosphocholaoate antisense oligodeoxynucleotides capable of  
 RT inhibiting Hepatitis C virus gene expression: In vitro translation  
 RT assay."  
 RL J. Blochem. 118:1199-1204(1995).  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; D45172; BAA08120.1; -  
 DR PIR; AG1196; AG1196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00329; P00329.  
 DR HSSP; P26663; IUXP.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_tyrp\_in.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR001490; HCV NS4A.  
 DR InterPro; IPR001490; HCV NS4B.  
 DR InterPro; IPR002868; HCV NS5A.  
 DR InterPro; IPR002166; HCV NS5B.  
 DR InterPro; IPR001650; HCV RdRp.  
 DR InterPro; IPR004109; Peptidase C9.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_BS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVIR.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326880 MW; EED840E6A050E766 CRC64;

Query Match 96.0%; Score 1970; DB 12; Length 3010;  
 Best Local Similarity 94.1%; Pred. No. 1.2e-155;  
 Matches 370; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 1 MASCGAVFGLALTLTSPYKVLARLIMWLYLITRVEAHLQWIPPLNVRGDPAL 60  
 DB 814 MASCGAVFGLVLLTSPYKVLARLIMWLYLITRVEAHLQWIPPLNVRGDPAL 873  
 QY 61 ILLTCAVHPELFDITKLLALFGLPLVLOAGITVVPVFAQGLIRACMLVKAAGHY 120  
 DB 874 ILLTCAVHPELFDITKLLALFGLPLVLOAGITVVPVFAQGLIRACMLVKAAGHY 933  
 QY 121 VGMAMKLAALGTVVVDHLPLODMAHAGLRLAVANPEVFSMEVKIITWGAADTAAC 180  
 DB 934 VGMAMKLAALGTVVVDHLPLODMAHAGLRLAVANPEVFSMEVKIITWGAADTAAC 993  
 QY 994 GIIIGLPVSARKEILLGPADSLGGQWRLLATITVSOQTRGLCGITISLGRDKN 1053  
 DB 181 GIIIGLPVSARKEILLGPADSLGGQWRLLATITVSOQTRGLCGITISLGRDKN 240  
 QY 241 QVEGEVQVSTAFATGCVTGVHGAQAKTLAGKPIITQMTNTVDODLVGMOS 300  
 DB 1054 QVEGEVQVSTAFATGCVTGVHGAQAKTLAGKPIITQMTNTVDODLVGMOS 1113  
 QY 301 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCP 360  
 DB 1114 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCP 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 9  
 Q81755  
 ID Q81755 PRELIMINARY; PRT; 1186 AA.  
 AC Q81755;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
 DE Polyprotein (fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9108850; PubMed=2175903;  
 RA Kato N., Hijioka K., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.,  
 RT "Molecular cloning of the human hepatitis C virus genome from Japanese  
 RT patients with non-A, non-B hepatitis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; D45172; BAA08120.1; -  
 DR PIR; AG1196; AG1196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00329; P00329.  
 DR HSSP; P26663; IUXP.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_tyrp\_in.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR001490; HCV NS4A.  
 DR InterPro; IPR001490; HCV NS4B.  
 DR InterPro; IPR002868; HCV NS5A.  
 DR InterPro; IPR002166; HCV NS5B.  
 DR InterPro; IPR001650; HCV RdRp.  
 DR InterPro; IPR004109; Peptidase C9.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_BS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVIR.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.

RT and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704 (1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91140698; Pubmed=1847440;  
 RA Takamizawa A., Mori C., Manabe S., Murakami S., Fujita J., Onishi E.,  
 RA Andoh T., Yoshida I., Okayama H.,  
 RT "The structure and organization of the Hepatitis C virus genome  
 RT isolated from human carriers.";  
 RL J. Virol. 65:1105-1113 (1991).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91172826; Pubmed=1848704;  
 RA Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C.,  
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.,  
 RA Bradley D.W., Kuo G., Houghton M.,  
 RT "Genetic organization and diversity of the hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455 (1991).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230206; Pubmed=1314449;  
 RA Chen P., Lin M., Tai K., Liu P., Lin C., Chen D.,  
 RA "The Taiwanese hepatitis C virus genome: Sequence determination and  
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113 (1992).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230232; Pubmed=1314459;  
 RA Okamoto H., Kurei K., Okada S., Yamamoto K., Iizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda F., Mishiro S.,  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: Comparative study of four distinct  
 RT genotypes.";  
 RL Virology 188:331-341 (1992).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93323208; Pubmed=8392606;  
 RA Hijikata M., Mizushima H., Akagi T., Mori S., Kakinuchi N., Kato N.,  
 RA Tanaka T., Kimura K., Shimotohno K.,  
 RT "Two distinct proteinase activities required for the processing of a  
 RT putative nonstructural precursor protein of hepatitis C virus.";  
 RL J. Virol. 67:4665-4675 (1993).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA Hijikata M.,  
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D11397; BAA20975.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; PS0329; PS0329.  
 DR PDB; 1DXF; 28-MAR-02.  
 DR GO; GO:0005524; P:ATP binding; IEA.  
 DR GO; GO:0008026; P:ATP dependent helicase activity; IEA.  
 DR GO; GO:0016787; F:Hydrolase activity; IEA.  
 DR GO; GO:0003676; F:Nucleic acid binding; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019087; P:Viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_Trypsin.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR000745; HCV NS4A.  
 DR InterPro; IPR001490; HCV NS4B.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR Pfam; PF01538; HCV NS2.1.  
 DR Pfam; PF02907; HCV NS3.1.  
 DR Pfam; PF01006; HCV NS4a.1.  
 DR Pfam; PF01001; HCV NS4b.1.  
 DR Pfam; PF00271; Helicase\_C.1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR ATP-binding; Helicase; Hydrolase.  
 FT NON TER 1  
 SQ SEQUENCE 1186 AA; 126280 MM; 34170478BA23729A CRC64;

Query Match 95.9%; Score 1968; DB 12; Length 1186;  
 Best Local Similarity 94.4%; Pred. No. 5.2e-156;  
 Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MAASGCAVFIGLALLTLSPYKYLARLIMWLQYLITRYEAHLQVWIPPLANRGSDAI 60  
 DB 92 MAASGCAVFIGLALLTLSPYKYLARLIMWLQYLITRYEAHLQVWIPPLANRGSDAI 151  
 QY 61 ILTCAVPELIPDITKLLALISPLVNLQAGIKVYFPAQGLIACMLVRRAGGHY 120  
 DB 152 ILTCAVPELIPDITKLLALISPLVNLQAGIKVYFPAQGLIACMLVRRAGGHY 211  
 QY 121 VQMAFMKALATGTYYVDHLTPLQDMAHAGRDIAVAVEPIFSDMEVKIITMGADTAAC 180  
 DB 212 VQMAFMKALATGTYYVDHLTPLQDMAHAGRDIAVAVEPIFSDMEVKIITMGADTAAC 271  
 QY 181 GDIISGLPVSARREKELLGPDNFEQGWRLAPITAYSQQTRGLICITTSITGSDKN 240  
 DB 272 GDIISGLPVSARREKELLGPDNFEQGWRLAPITAYSQQTRGLICITTSITGSDKN 331  
 QY 241 QVEGEVQVSTATQSFATCTNGVCMTEFGAGSKTLAGPKPTOMYTNVDDLVGMOA 300  
 DB 332 QVDSGVQVSTATQSFATCTNGVCMTEFGAGSKTLAGPKPTOMYTNVDDLVGMPA 391  
 QY 301 PGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRSGLSPREVSYLKSSGGPILCP 360  
 DB 392 PGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRSGLSPREVSYLKSSGGPILCP 451  
 QY 361 GHANGIFPAANTCGVAKAUNDFIVESMETMTR 393  
 DB 452 GHANGIFPAANTCGVAKAUNDFIVESMETMTR 484  
 RESULT 10  
 ID 081817 PRELIMINARY; PRT; 2284 AA.  
 AC 081817;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyprotein precursor (Genome polypeptide).  
 OS Hepatitis C virus type 2.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=40271;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9406484; Pubmed=7504283;  
 RA Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,  
 RA Akagi T., Kimura K., Shimotohno K.,  
 RT "Proteolytic processing and membrane association of putative  
 RT nonstructural proteins of hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777 (1993).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94333910; Pubmed=8056334;  
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.,  
 RT "Identification of the domain required for trans-cleavage activity of  
 RT hepatitis C viral serine proteinase.";  
 RL Gene 145:215-219 (1994).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95056078; Pubmed=7966638;  
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.,  
 RT "Hepatitis C virus polyprotein processing: kinetics and mutagenic  
 RT analysis of serine proteinase-dependent cleavage.";  
 RL J. Virol. 68:8418-8422 (1994).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95156583; Pubmed=7853491;  
 RA Tanji Y., Hijikata M., Satoh S., Kaneko T., Shimotohno K.,  
 RT "Hepatitis C virus-encoded nonstructural protein NS4A has versatile



Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
1	95.9%	371	1968	12	2284	0	0
8	94.4%	371	1261	15	2284	0	0
61	94.4%	371	1261	15	2284	0	0
148	94.4%	371	1261	15	2284	0	0
121	94.4%	371	1261	15	2284	0	0
208	94.4%	371	1261	15	2284	0	0
181	94.4%	371	1261	15	2284	0	0
268	94.4%	371	1261	15	2284	0	0
241	94.4%	371	1261	15	2284	0	0
328	94.4%	371	1261	15	2284	0	0

```

QY      30.1 PEARSMTCTCGSSDPLVYVTHAAVIVYRRGRSGRLSPREVSYLKSSGGPILCP 36
Db      388 PPARMTCTCGSSDPLVYVTHAAVIVYRRGRSGRLSPREVSYLKSSGGPILCP 447

QY      361 GHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
Db      448 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 480

RESULT 11
P89966 PRELIMINARY; PRT; 3010 AA.
P89966
AC P89966;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE RNA for polyprotein (Genome polyprotein).
OS Hepatitis C virus.
OC viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=type 1b;
RA Tanaka T.;
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=type 1b;
RA TANAKA T.;
RT "MORE ";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC
CC EMBL; D89872; EMBL4035.1; -.
DR DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; P00804; P00804.
DR PIR; P50329; P50329.
DR HSSP; P26663; 1JXP.
DR DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003668; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; F:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR DR InterPro; IPR009003; Cys_Ser_typsin.
DR DR InterPro; IPR001410; DEAD.
DR DR InterPro; IPR002522; HCV capsid.
DR DR InterPro; IPR002521; HCV core.
DR DR InterPro; IPR002519; HCV_enu.
DR DR InterPro; IPR002518; HCV_NS1.
DR DR InterPro; IPR002518; HCV_NS2.
DR DR InterPro; IPR00745; HCV_NS4.
DR DR InterPro; IPR001490; HCV_NS4b.
DR DR InterPro; IPR002868; HCV_NS5a.
DR DR InterPro; IPR002166; HCV_NS5b.
DR DR InterPro; IPR001650; Helicase_C.
DR DR InterPro; IPR004109; peptidase_C29.
DR DR InterPro; IPR007095; RNA_pol_DS_P5.
DR DR InterPro; IPR007094; RNA_pol_P5vir.
DR Pfam; PF01543; HCV_capsid.1.

```

DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR Pfam: PF00487; DEXDC; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327023 MW; E075BD9CED801261 CRC64;

Query Match 95.9%; Score 1968; DB 12; Length 3010;  
 Best Local Similarity 94.4%; Pred. No. 1.7e-155;  
 Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

DR 1 MAASCGAVFGLALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGDAI 60  
 DB 814 MAASCGAVFGLALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGDAI 873  
 QY 61 ILITCAVHPELIPITKLLAIPGLMWLQGITKVPYFAQGLIRACMLVRKAAGHY 120  
 DB 874 ILITCAVHPELIPITKLLAIPGLMWLQGITKVPYFAQGLIRACMLVRKAAGHY 933  
 QY 121 VQMAFMKLAALGTGYVDHLLPQDMAHAGRLDAVAVEPVISDMEVKIITMGADTAAC 180  
 DB 934 VQMAFMKLAALGTGYVDHLLPQDMAHAGRLDAVAVEPVISDMEVKIITMGADTAAC 993  
 QY 181 GDIISGLPVSARRGRIILGPADNFEQGRLLAPITAYSQOTRGLGCIITSLTRDKN 240  
 DB 994 GDIISGLPVSARRGRIILGPADNFEQGRLLAPITAYSQOTRGLGCIITSLTRDKN 1053  
 QY 241 QVSEGVQVSTATQSFATCVNGVCTVFEGASSKTLAGKXPIITQMTNVDLGVNKA 300  
 DB 1054 QVSEGVQVSTATQSFATCVNGVCTVFEGASSKTLAGKXPIITQMTNVDLGVNKA 1113  
 QY 301 PPGARSTPCTCGSSDLYLTRHADVPVARRDSSGSLSPRVLYLKGSSGFPILCS 360  
 DB 1114 PPGARSTPCTCGSSDLYLTRHADVPVARRDSSGSLSPRVLYLKGSSGFPILCS 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 1206

## RESULT 12

ID 09J3G6 PRELIMINARY; PRT; 3010 AA.  
 AC 09J3G6;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ND26;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC LIPOPROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; AF207767; AAF65957.1; -

DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: P00254; P00254.  
 DR PIR: P00329; P00329.  
 DR HSRP: P26663; IUXP.  
 DR MEROPS: S29.002; -.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transporter activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006118; F: electron transport; IEA.  
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser tyrosin.  
 DR InterPro: IPR000345; CytC\_heme\_35.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA pol DS PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327165 MW; 74FAB6B80F24837B CRC64;

Query Match 95.9%; Score 1968; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 1.7e-155;  
 Matches 372; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGDAI 60  
 DB 814 MAASCGAVFGLALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGDAI 873  
 QY 61 ILITCAVHPELIPITKLLAIPGLMWLQGITKVPYFAQGLIRACMLVRKAAGHY 120  
 DB 874 ILITCAVHPELIPITKLLAIPGLMWLQGITKVPYFAQGLIRACMLVRKAAGHY 933  
 QY 121 VQMAFMKLAALGTGYVDHLLPQDMAHAGRLDAVAVEPVISDMEVKIITMGADTAAC 180  
 DB 934 VQMAFMKLAALGTGYVDHLLPQDMAHAGRLDAVAVEPVISDMEVKIITMGADTAAC 993  
 QY 181 GDIISGLPVSARRGRIILGPADNFEQGRLLAPITAYSQOTRGLGCIITSLTRDKN 240

```

Db 994 GDIIILG:PVARSARGREILIGPADSFEGQWRLLAPITAYSQTRGLGCIITSLTGRDXN 1053
Qy 241 QVEGEVQVNSTAQTQSPFLATCVNGVCMVTFVHGASKTLGKSGITQMTYNVDQDLVGMQA 300
Db 1054 QVEGEVQVNSTAQTQSPFLATCVNGVCMVTFVHGASKTLGKSGITQMTYNVDQDLVGMQA 1113
Qy 301 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPRVSYLKSGSGPILCPS 360
Db 1114 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPRVSYLKSGSGPILCPS 1173
Qy 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 393
Db 1174 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 13
Q99AU2 PRELIMINARY; PRT; 3010 AA.
AC Q99AU2;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Genome polypeptide.
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN NCBI_Taxid=31647;
RP SEQUENCE FROM N.A.
RA STRAIN=chimera of HCV-BK;
RA Thomson M., Nascent M., Gonzales S., Murthy K., Rehmann B.,
RA Liang J.;
RT "Analyses of viral sequences and virus-specific immune responses
RT during serial passage of an infectious hepatitis C virus serotype 1b
RT clone in chimpanzees";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL: AF333324; AAK08509.1; -.
DR PIR: A61196; A61196.
DR PIR: P00246; P00246.
DR PIR: P00804; P00804.
DR PIR: P03229; P03229.
DR HSSP: P26663; INS3.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F: electron transporter activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0016740; F: transferase activity; IEA.
DR GO: GO:0006118; P: electron transport; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P: transcription; IEA.
DR GO: GO:0019079; P: viral genome replication; IEA.
DR GO: GO:0019087; P: viral transformation; IEA.
DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR00522; HCV_capsid.
DR InterPro: IPR00521; HCV_core.
DR InterPro: IPR00519; HCV env.
DR InterPro: IPR00531; HCV NS1.
DR InterPro: IPR00518; HCV NS2.
DR InterPro: IPR00745; HCV NS4a.
DR InterPro: IPR001490; HCV_NS4b.

```

```

DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.
DR InterPro: IPR004109; Peptidase C29.
DR InterPro: IPR007095; RNA_pol_DS_P5.
DR InterPro: IPR007094; RNA_pol_PSVit.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV NS1; 1.
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02807; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV NS1; 1.
DR SMART: SM00487; DEXDC_1.
DR PROSITE: PS00190; CYTOCHROME C; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327007 MW; 053B9A653B0AB335 CRC64;

Query Match 95.8%; Score 1967; DB 12; Length 3010;
Best Local Similarity 93.9%; Pred. No. 2.1e-155;
Matches 369; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MAASCGAVFIGLALITLSPYKVLARLIWLYQLITRVEAHQVWIPLVNAGRDAL 60
Db 814 MAASCGAVFGLVLTLSFYKVFARLIWLYQLITRVEAHQVWIPLVNAGRDAL 873
Qy 61 ILTCAVHPELFDITKLLAFSGPLMLOAGITVFPYVAGGILRACMVRKAGGHY 120
Db 874 ILTCAVHPELFDITKLLAFSGPLMLOAGITVFPYVAGGILRACMVRKAGGHY 933
Qy 121 VQMAFMKLAALGTGVYVYHLPLOMAHAGLRLVAVEPVYFSMEVKIITMGADTAAC 180
Db 934 VQMAFMKLAALGTGVYVYHLPLOMAHAGLRLVAVEPVYFSMEVKIITMGADTAAC 993
Qy 181 GDIIGLPSVARSARGEIILGPADNFEQGWRLAPITAYSQTRGLGCIITSLTGRDXN 240
Db 994 GDIIGLPSVARSARGEIILGPADNFEQGWRLAPITAYSQTRGLGCIITSLTGRDXN 1053
Qy 241 QVEGEVQVNSTAQTQSPFLATCVNGVCMVTFVHGASKTLGKSGITQMTYNVDQDLVGMQA 300
Db 1054 QVEGEVQVNSTAQTQSPFLATCVNGVCMVTFVHGASKTLGKSGITQMTYNVDQDLVGMQA 1113
Qy 301 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPRVSYLKSGSGPILCPS 360
Db 1114 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPRVSYLKSGSGPILCPS 1173
Qy 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 393
Db 1174 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 14
Q99TD7 PRELIMINARY; PRT; 3010 AA.
AC Q99TD7;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Genome polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN NCBI_Taxid=11103;
RP SEQUENCE FROM N.A.
RA STRAIN=HCV217;
RA Takahashi K., Ohta Y., Kanai K., Maruo H., Baba K., Hijioka M.,
RA Mishiro S.;

```

"Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 'progression score' revisited." RT  
 Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases. RL  
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A CC  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: CC  
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF CC  
 PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: AB049100; BAB1813.1; -.  
 DR PIR: A61196; A61196.  
 DR PIR: PQ0804; PQ0804.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P26663; IUXP.  
 DR MEROPS: S29.002; -.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transporter activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006118; F: electron transport; IEA.  
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; F: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; Cyt\_heme\_35.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS4d.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS4d; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR Pfam: PF00602; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR KEGG: K00001; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transmembrane. K  
 SEQUENCE 3010 AA; 326909 MW; 5505C62EB0DA0519 CRC64; S0

Query Match 95.8%; Score 1967; DB 12; Length 3010;  
 Best Local Similarity 94.4%; Pred. No. 2.1e-155;  
 Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKYLARLIMWLYLTRVEAHLYQWIPPLNVRGGDAI 60  
 DB 814 VAASCGAVFGLALTLSPYKYLARLIMWLYLTRVEAHLYQWIPPLNVRGGDAI 873  
 QY 61 ILTCAVHPELIPITLLAIFGLPLVLAQGITKPYFPRACGLRACMLVKKAGGY 120

DB 874 ILTCAVHPELIPITLLAIFGLPLVLAQGITRVEYFVPAQGLIRACMLVKKAGGY 933  
 QY 121 VQAMFKLAALTYGVYDHLPLQDMNAAGRDLAVVEPIFSDMEVKIITMGADTAAC 180  
 DB 934 IQMLVLAALTYGVYDHLPLQDMNAAGRDLAVVEPIFSDMEVKIITMGADTAAC 993  
 QY 181 GDIISGLVPSARRREILGPADNFEQGMWLLAPITAYSQQTRGLIGCIITSLTGDKN 240  
 DB 994 GDIIISGLVPSARRREILGPADNFEQGMWLLAPITAYSQQTRGLIGCIITSLTGDKN 1053  
 QY 241 QVEGEVQVSTATOSFLATCNVGCWTFHGAGSKTLAGPKPTQMTYTNVDDLVGMOA 300  
 DB 1054 QVEGEVQVSTATOSFLATCNVGCWTFHGAGSKTLAGPKPTQMTYTNVDDLVGMOA 1113  
 QY 301 PPGARSMPTCTCGSSDLYLTRHADVIPIVRRGDSRSLSPRPVSYLKSSGGPLLCP 360  
 DB 1114 PPGARSMPTCTCGSSDLYLTRHADVIPIVRRGDSRSLSPRPVSYLKSSGGPLLCP 1173  
 QY 361 GHANGIFPAANCTGVAKADVIFVESMETTR 393  
 DB 1174 GHANGIFPAANCTGVAKADVIFVESMETTR 1206

RESULT 15  
 ID Q9QIX6 PRELIMINARY; PRT; 3010 AA.  
 AC Q9QIX6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD8-1;  
 RX MEDLINE=20013325; PubMed=10544098;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,  
 RA Tazawa J.I., Izumi N., Matsumo F., Sato C.;  
 RT "Time-related changes in full-length hepatitis C virus and hepatitis  
 RT activity.";  
 RL Virology 263:244-253 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD8-1;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,  
 RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Matsumo F., Sato C.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AF15059; AAD56194.1; -.  
 DR PIR: A61196; A61196.  
 DR PIR: PQ0246; PQ0246.  
 DR PIR: PQ0254; PQ0254.  
 DR PIR: PQ0804; PQ0804.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P26663; IUXP.  
 DR MEROPS: S29.002; -.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_cryptin.  
 DR InterPro; IPR004103; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NSI.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF002271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR PRODOM; PD186062; DEXdc; 1.  
 DR SMART; SM00487; DEXdc; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327466 MW; 4613744EC4DA013 CRC64;

Query Match 95.8%; Score 1967; DB 12; Length 3010;  
 Best Local Similarity 94.4%; Pred. No. 2,1e-155;

Matches 371; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 1 MAASCGAVFIQALTLSPYKTLARLIMLOYLITRVEAHLOVWIPPLNVRGGRDAI 60  
 |||||  
 DB 814 MAASCGAVFVGLAFLTSPYKFLARLIMLOYLITRVEAHLOVWIPPLNVRGGRDAI 873  
 |||||  
 QY 61 ILTCAVHPELIPDITKLLAIFGPMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHY 120  
 |||||  
 DB 874 IILMCVHPELIPDITKLLAIFGPMVLQAGITRVPYFVRAHGLIRACMLVRKVGHY 933  
 |||||  
 QY 121 VQMAFMTLALTGYVVDHLTPIQDMAHGLRDLAFAVEVPIPSDMEVKIITWGADTAAC 180  
 |||||  
 DB 934 VQMAFVTLALTGYVVDHLAPIOHMAHSGRLDLAFAVEVPIPSDMEVKIITWGADTAAC 993  
 |||||  
 QY 181 GDITSGLPVSARRGREILLGPADNFEQGWRLAPITAYSQOTRGLIGCIITSLTGRDKN 240  
 |||||  
 DB 994 GDITLGLPVSARRGREILLGPADNFEQGWRLAPITAYSQOTRGLIGCIITSLTGRDKN 1053  
 |||||  
 QY 241 QVEGEVQVYATATQSFATCTGAGVCTVFRGAGSKITAGKGPITQYTYTNDODLVGMQA 300  
 |||||  
 DB 1054 QVEGEVQVYATATQSFATCTGAGVCTVFRGAGSKITAGKGPITQYTYTNDODLVGMQA 1113  
 |||||  
 QY 301 PGARSMTPCTCGSSDLYLVTRHADVI PVRARGDSRGLSPRPVSYLKSSGGPLLCPS 360  
 |||||  
 DB 1114 PGARSMTPCTCGSSDLYLVTRHADVI PVRARGDSRGLSPRPVSYLKSSGGPLLCPS 1173  
 |||||  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPESMETTMR 393  
 |||||  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFIPESMETTMR 1206  
 |||||

Search completed: May 6, 2004, 09:35:46  
 Job time : 37.7566 secs



Fri May 7 13:37:00 2004

us-10-650-585-11.fai

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004. CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 15.4246 Seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MAASCGAVFGLALTLTSP.....RGVAKXVDPIPVESMETTR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCUTS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1951	95.0	2201	4 US-09-539-601-6	Sequence 6, Appl
2	1951	95.0	2201	4 US-09-539-601-15	Sequence 15, Appl
3	1951	95.0	3010	4 US-09-539-601-3	Sequence 3, Appl
4	1951	95.0	3010	4 US-09-539-601-21	Sequence 21, Appl
5	1951	95.0	3010	4 US-09-539-601-27	Sequence 27, Appl
6	1946	94.8	1692	3 US-09-263-933-4	Sequence 4, Appl
7	1946	94.8	1692	3 US-09-263-933-2	Sequence 2, Appl
8	1946	94.8	2307	3 US-09-263-933-2	Sequence 2, Appl
9	1946	94.8	2307	3 US-09-263-933-2	Sequence 2, Appl
10	1944	94.7	3010	4 US-09-539-601-13	Sequence 3, Appl
11	1943	94.6	1692	3 US-09-263-933-11	Sequence 11, Appl
12	1943	94.6	1692	3 US-09-263-933-9	Sequence 9, Appl
13	1943	94.6	2307	3 US-09-263-933-9	Sequence 9, Appl
14	1943	94.6	2307	3 US-09-263-933-18	Sequence 18, Appl
15	1934	94.2	1692	3 US-09-263-933-18	Sequence 18, Appl
16	1934	94.2	2307	3 US-09-263-933-16	Sequence 16, Appl
17	1934	94.2	2307	3 US-09-263-933-16	Sequence 16, Appl
18	1934	94.2	2307	3 US-09-263-933-16	Sequence 16, Appl
19	1928	93.9	3010	3 US-09-539-601-16	Sequence 3, Appl
20	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl
21	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl
22	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl
23	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl
24	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl
25	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl
26	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl
27	1888	92.0	2013	1 US-08-324-977-12	Sequence 12, Appl

28	1888	92.0	2620	3 US-09-315-850-32	Sequence 32, Appl
29	1888	92.0	2621	1 US-08-324-977-36	Sequence 36, Appl
30	1888	92.0	2621	1 US-08-324-977-36	Sequence 36, Appl
31	1888	92.0	2621	1 US-08-324-977-36	Sequence 36, Appl
32	1888	92.0	2621	1 US-08-324-977-36	Sequence 36, Appl
33	1888	92.0	3010	1 US-08-324-977-2	Sequence 2, Appl
34	1888	92.0	3010	1 US-08-324-977-14	Sequence 14, Appl
35	1888	92.0	3010	1 US-08-324-977-14	Sequence 14, Appl
36	1888	92.0	3010	1 US-08-324-977-14	Sequence 14, Appl
37	1888	92.0	3010	1 US-08-324-977-14	Sequence 14, Appl
38	1888	92.0	3010	1 US-08-324-977-14	Sequence 14, Appl
39	1888	92.0	3010	1 US-08-324-977-14	Sequence 14, Appl
40	1888	92.0	3010	1 US-08-324-977-14	Sequence 14, Appl
41	1767	86.1	1648	5 US-08-188-281B-12	Sequence 12, Appl
42	1767	86.1	1648	5 US-08-188-281B-12	Sequence 12, Appl
43	1767	86.1	1648	5 US-08-188-281B-12	Sequence 12, Appl
44	1767	86.1	1648	5 US-08-188-281B-12	Sequence 12, Appl
45	1767	86.1	3011	1 US-08-453-552-1	Sequence 1, Appl

# ALIGNMENTS

RESULT 1  
US-09-539-601-6  
; Sequence 6, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlag, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; EARLIER FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-6

Query Match 95.0%; Score 1951; DB 4; Length 2201;  
Best Local Similarity 93.1%; Pred. No. 8.8e-187;  
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY	1	MAASCGAVFGLALTLTSPYKVLARLIMWLQYLITRVEAHLQVMIPLNVRGRDAI	60
DB	5	MAASCGAVFGLALTLTSPHYKFLARLIMWLQYFIRAEHLQVMIPLNVRGRDAV	64
QY	61	IIITCAVPEELFDITKLLALFGLPLMTQAGITVPPVRAQGLIRACMLYRKAAGHY	120
DB	65	IIITCAVPEELFDITKLLALFGLPLMTQAGITVPPVRAQGLIRACMLYRKAAGHY	124
QY	121	VOMAFKLLALGTGVYVDHLFPLQDMAGLRDLAAVEPVIFSDMEVKIITWGAOTAC	180
DB	125	VOMAFKLLALGTGVYVDHLFPLQDMAGLRDLAAVEPVIFSDMEVKIITWGAOTAC	184
QY	181	GDIISGLPVASRGRBILIPADNFEQGRLLAPITAYSOQTRGLGCIITSLGRDN	240
DB	185	GDIISGLPVASRGRBILIPADNFEQGRLLAPITAYSOQTRGLGCIITSLGRDN	244
QY	241	QVEGEVQVSTASQSFATCVNGVCMTHFGAGSKTLGPKPITOMYTNVDDLVGQA	300
DB	245	QVEGEVQVSTASQSFATCVNGVCMTHFGAGSKTLGPKPITOMYTNVDDLVGQA	304
QY	301	PPGARSMTCTCGSSDLYLVTRHADVIPYRRRGRSGLISRPVSYLKGSSGGLLCP	360
DB	305	PPGARSMTCTCGSSDLYLVTRHADVIPYRRRGRSGLISRPVSYLKGSSGGLLCP	364
QY	361	GHAAGFRAAVCTRGVAKXVDPIPVESMETTR	393



Db 365 GHAVGIFRAAVCTRGVAKAVDFVVESEMETTMR 397

RESULT 2  
US-09-539-601-15  
Sequence 15, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartenschlager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 15  
LENGTH: 2201  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-15

Query Match 95.0%; Score 1951; DB 4; Length 2201;  
Best Local Similarity 93.1%; Pred. No. 8.8e-187;  
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Db 1 MAASCGAVFIGALITLSPYKVLARLIWLOYLITRVEAHLOVWIPPLNVRGGDAI 60  
5 MAASCGAVFIGALITLSPYKVLARLIWLOYLITRVEAHLOVWIPPLNVRGGDAV 64  
QY 61 ILTCAVHPELIFDITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
Db 65 ILTCAHPELIFITITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 124  
QY 121 VQAFMKLAALTGYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAAC 180  
Db 125 VQAFMKLAALTGYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAAC 184  
QY 181 GDITSGLPVSARRREILHGPADNFEQGWRLAPITAVSQQTRGLIGCIITSLTGRDN 240  
Db 185 GDITSGLPVSARRREILHGPADNFEQGWRLAPITAVSQQTRGLIGCIITSLTGRDN 244  
QY 241 QVEGEVQVSTATQSFATCVNGVCFHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Db 245 QVEGEVQVSTATQSFATCVNGVCFHAGSKTLAGKGPITOMYTNVDDLVGMQA 304  
QY 301 PPGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRVSYLKSSGGPLLCPS 360  
Db 305 PPGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRVSYLKSSGGPLLCPS 364  
QY 361 GHAVGIFRAAVCTRGVAKAVDFVVESEMETTMR 393  
Db 365 GHAVGIFRAAVCTRGVAKAVDFVVESEMETTMR 397

RESULT 3  
US-09-539-601-3  
Sequence 3, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartenschlager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 3  
LENGTH: 3010  
TYPE: PRT

ORGANISM: Hepatitis C virus  
US-09-539-601-3

Query Match 95.0%; Score 1951; DB 4; Length 3010;  
Best Local Similarity 93.1%; Pred. No. 1.4e-186;  
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFIGALITLSPYKVLARLIWLOYLITRVEAHLOVWIPPLNVRGGDAI 60  
Db 814 MAASCGAVFIGALITLSPYKVLARLIWLOYLITRVEAHLOVWIPPLNVRGGDAV 873  
QY 61 ILTCAVHPELIFDITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
Db 874 ILTCAHPELIFITITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 933  
QY 121 VQAFMKLAALTGYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAAC 180  
Db 934 VQAFMKLAALTGYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAAC 993  
QY 181 GDITSGLPVSARRREILHGPADNFEQGWRLAPITAVSQQTRGLIGCIITSLTGRDN 240  
Db 994 GDITSGLPVSARRREILHGPADNFEQGWRLAPITAVSQQTRGLIGCIITSLTGRDN 1053  
QY 241 QVEGEVQVSTATQSFATCVNGVCFHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Db 1054 QVEGEVQVSTATQSFATCVNGVCFHAGSKTLAGKGPITOMYTNVDDLVGMQA 1113  
QY 301 PPGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRVSYLKSSGGPLLCPS 360  
Db 1114 PPGARSMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGLSPRVSYLKSSGGPLLCPS 1173  
QY 361 GHAVGIFRAAVCTRGVAKAVDFVVESEMETTMR 393  
Db 1174 GHAVGIFRAAVCTRGVAKAVDFVVESEMETTMR 1206

RESULT 4  
US-09-539-601-21  
Sequence 21, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartenschlager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 21  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-21

Query Match 95.0%; Score 1951; DB 4; Length 3010;  
Best Local Similarity 93.1%; Pred. No. 1.4e-186;  
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFIGALITLSPYKVLARLIWLOYLITRVEAHLOVWIPPLNVRGGDAI 60  
Db 814 MAASCGAVFIGALITLSPYKVLARLIWLOYLITRVEAHLOVWIPPLNVRGGDAV 873  
QY 61 ILTCAVHPELIFDITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
Db 874 ILTCAHPELIFITITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 933  
QY 121 VQAFMKLAALTGYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAAC 180  
Db 934 VQAFMKLAALTGYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAAC 993  
QY 181 GDITSGLPVSARRREILHGPADNFEQGWRLAPITAVSQQTRGLIGCIITSLTGRDN 240

DB 994 GDIILGLPVSARGRGRIHIGPADSLEGQGWRLAPITAVSQOTRGLGCIITSLTGRDKN 1053  
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFPHGAGSKTLGPKGPITQMTYTNVDQDLVGMQA 300  
DB 1054 QVEGEVQVSTATQSFATCVNGVCMVTFPHGAGSKTLGPKGPITQMTYTNVDQDLVGMQA 1113  
QY 301 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRRGDSRGSILSPRPVSYLKSGSGPILCP 360  
DB 1114 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRRGDSRGSILSPRPVSYLKSGSGPILCP 1173  
QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTR 393  
DB 1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTR 1206

## RESULT 5

US-09-539-601-27  
Sequence 27, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartenschlager, Ralf FM  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
EARLIER FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 27  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-27

Query Match 95.0%; Score 1951; DB 4; Length 3010;  
Best Local Similarity 93.1%; Pred. No. 1,4e-186;  
Matches 365; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQVWIPPLNARGSDAI 60  
DB 814 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQVWIPPLNARGSDAI 873  
QY 61 ILTCAVHPELFDITKLLAIFGPIVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 120  
DB 874 ILTCAVHPELFDITKLLAIFGPIVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 933  
QY 121 VQMAFMKLAALTGIVYVDHLTPLQDMAHAGRLDAVAEPIFSDMEVKIITWGADTAA 180  
DB 934 VQMAFMKLAALTGIVYVDHLTPLQDMAHAGRLDAVAEPIFSDMEVKIITWGADTAA 993  
QY 181 GDITSGIPVARSRRREILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 240  
DB 994 GDITSGIPVARSRRREILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 1053  
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFPHGAGSKTLGPKGPITQMTYTNVDQDLVGMQA 300  
DB 1054 QVEGEVQVSTATQSFATCVNGVCMVTFPHGAGSKTLGPKGPITQMTYTNVDQDLVGMQA 1113  
QY 301 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRRGDSRGSILSPRPVSYLKSGSGPILCP 360  
DB 1114 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRRGDSRGSILSPRPVSYLKSGSGPILCP 1173  
QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTR 393  
DB 1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTR 1206

RESULT 6  
US-09-263-933-4  
Sequence 4, Application US/09263933  
Patent No. 6280940

GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
EARLIER FILING DATE: 1998-03-08  
EARLIER APPLICATION NUMBER: 09/129,611  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-4

Query Match 94.8%; Score 1946; DB 3; Length 1692;  
Best Local Similarity 92.9%; Pred. No. 1,9e-186;  
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQVWIPPLNARGSDAI 60  
DB 93 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQVWIPPLNARGSDAI 152  
QY 61 ILTCAVHPELFDITKLLAIFGPIVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 120  
DB 153 ILTCAVHPELFDITKLLAIFGPIVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 212  
QY 121 VQMAFMKLAALTGIVYVDHLTPLQDMAHAGRLDAVAEPIFSDMEVKIITWGADTAA 180  
DB 213 VQMAFMKLAALTGIVYVDHLTPLQDMAHAGRLDAVAEPIFSDMEVKIITWGADTAA 272  
QY 181 GDITSGIPVARSRRREILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 240  
DB 273 GDITSGIPVARSRRREILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 332  
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFPHGAGSKTLGPKGPITQMTYTNVDQDLVGMQA 300  
DB 333 QVEGEVQVSTATQSFATCVNGVCMVTFPHGAGSKTLGPKGPITQMTYTNVDQDLVGMQA 392  
QY 301 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRRGDSRGSILSPRPVSYLKSGSGPILCP 360  
DB 393 PGARSMTPTCTGSSDLYLVTRHADVIPIVRRRGDSRGSILSPRPVSYLKSGSGPILCP 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTR 485

RESULT 7  
US-09-919-901-4  
Sequence 4, Application US/09919901  
Patent No. 6539738  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
EARLIER FILING DATE: 2001-08-02  
EARLIER APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1692

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-4

Query Match 94.8%; Score 1946; DB 4; Length 1692;

Best Local Similarity 92.9%; Pred. No. 1.9e-186; Indels 0; Gaps 0;

Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLYLITRVEALQVWIPPLNVRGRDAI 60  
DB 93 MAASCGAVFGLALTLSPYKVLARLIWLYLITRVEALQVWIPPLNVRGRDAI 152  
QY 61 ILTCAVHPELIDITLKLALFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 120  
DB 153 ILMCAVHPELIDITLKLALFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 212  
QY 121 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 180  
DB 213 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 272  
QY 181 GDIISGLPVARSRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDXN 240  
DB 273 GDIISGLPVARSRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDXN 332  
QY 241 QVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAGPKGPIITQWYTNVDQDLVGMQA 300  
DB 333 QVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAGPKGPIITQWYTNVDQDLVGMQA 392  
QY 301 PPGARSMPTCTGSSDLYLTRHADVIYVRRRGRSGLSPRPVSYLKSSGGPILCP 360  
DB 393 PPGARSMPTCTGSSDLYLTRHADVIYVRRRGRSGLSPRPVSYLKSSGGPILCP 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 485

## RESULT 8

US-09-263-933-2

Sequence 2, Application US/09263933

Patent No. 6280940

GENERAL INFORMATION:

APPLICANT: Potts, Karen E.

APPLICANT: Jackson, Roberta L.

APPLICANT: Patrick, Amy K.

TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-0005A

CURRENT APPLICATION NUMBER: US/09/263,933

CURRENT FILING DATE: 1999-03-08

EARLIER APPLICATION NUMBER: 09/129,611

EARLIER FILING DATE: 1998-08-05

NUMBER OF SEQ ID NOS: 33

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2

LENGTH: 2307

TYPE: PRT

ORGANISM: Artificial Sequence

US-09-263-933-2

Query Match 94.8%; Score 1946; DB 3; Length 2307;

Best Local Similarity 92.9%; Pred. No. 3e-186;

Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLYLITRVEALQVWIPPLNVRGRDAI 60  
DB 185 MAASCGAVFGLALTLSPYKVLARLIWLYLITRVEALQVWIPPLNVRGRDAI 244  
QY 61 ILTCAVHPELIDITLKLALFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 120  
DB 245 ILMCAVHPELIDITLKLALFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 304

QY 121 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 180  
DB 305 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 364  
QY 181 GDIISGLPVARSRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDXN 240  
DB 365 GDIISGLPVARSRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDXN 424  
QY 241 QVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAGPKGPIITQWYTNVDQDLVGMQA 300  
DB 425 QVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAGPKGPIITQWYTNVDQDLVGMQA 484  
QY 301 PPGARSMPTCTGSSDLYLTRHADVIYVRRRGRSGLSPRPVSYLKSSGGPILCP 360  
DB 485 PPGARSMPTCTGSSDLYLTRHADVIYVRRRGRSGLSPRPVSYLKSSGGPILCP 544  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 577

## RESULT 9

US-09-919-901-2

Sequence 2, Application US/09919901

Patent No. 6599738

GENERAL INFORMATION:

APPLICANT: Potts, Karen E.

APPLICANT: Jackson, Roberta L.

APPLICANT: Patrick, Amy K.

TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-0005A

CURRENT APPLICATION NUMBER: US/09/919,901

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 09/263,933

PRIOR FILING DATE: 1999-02-08

PRIOR APPLICATION NUMBER: 09/129,611

PRIOR FILING DATE: 1998-08-05

NUMBER OF SEQ ID NOS: 33

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2

LENGTH: 2307

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: :

US-09-919-901-2

Query Match 94.8%; Score 1946; DB 4; Length 2307;

Best Local Similarity 92.9%; Pred. No. 3e-186;

Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLYLITRVEALQVWIPPLNVRGRDAI 60  
DB 185 MAASCGAVFGLALTLSPYKVLARLIWLYLITRVEALQVWIPPLNVRGRDAI 244  
QY 61 ILTCAVHPELIDITLKLALFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 120  
DB 245 ILMCAVHPELIDITLKLALFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 304  
QY 121 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 180  
DB 305 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 364  
QY 181 GDIISGLPVARSRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDXN 240  
DB 425 GDIISGLPVARSRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDXN 484  
QY 301 PPGARSMPTCTGSSDLYLTRHADVIYVRRRGRSGLSPRPVSYLKSSGGPILCP 360  
DB 485 PPGARSMPTCTGSSDLYLTRHADVIYVRRRGRSGLSPRPVSYLKSSGGPILCP 544  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 577

QY 301 PGARSMCTCGSSDLVLTTRHADVIPIVRRGDSRGSLLSPRVSYLKSGSGPILCPSS 360  
 DB 485 PPARSLTPTCTGSSDLVLTTRHADVIPIVRRGDSRGSLLSPRVSYLKSGSGPILCPSS 544  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 393  
 DB 545 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 577

## RESULT 10

US-09-539-601-33  
 / Sequence 33, Application US/09539601C  
 / Patent No. 6630343  
 / GENERAL INFORMATION:  
 / APPLICANT: Batten Schlager, Ralf FW  
 / TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 / FILE REFERENCE: all sequences  
 / CURRENT APPLICATION NUMBER: US/09/539,601C  
 / CURRENT FILING DATE: 2001-08-30  
 / EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 / EARLIER FILING DATE: 1999-04-03  
 / NUMBER OF SEQ ID NOS: 51  
 / SOFTWARE: PatentIn Ver. 2.1  
 / SEQ ID NO 33  
 / LENGTH: 3010  
 / TYPE: PRT  
 / ORGANISM: Hepatitis C virus  
 US-09-539-601-33

Query Match 94.7%; Score 1944; DB 4; Length 3010;  
 Best Local Similarity 92.3%; Pred. No. 7.1e-186;  
 Matches 365; Conservative 14; Mismatches 14; Indels 0; Gaps 0;

QY 1 MAASCGAVFIQGLALTLSPYKVLARLWMLQYLITRVEAHQWIPPLNVRGSDAI 60  
 DB 814 MAASCGAVFVGLVLTLSPIYKFLARLWMLQYFTTRAEHLQWIPPLNVRGSDAV 873  
 QY 61 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAGGHY 120  
 DB 874 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAGGHY 933  
 QY 121 VQAMFKLAALTGYVVDHLTPLODMAHAGLDLAVAVEPVFSDETKIITWGADTAAC 180  
 DB 934 VQAMFKLAALTGYVVDHLTPLODMAHAGLDLAVAVEPVFSDETKIITWGADTAAC 993  
 QY 181 GDIISGLFVSARRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 240  
 DB 994 GDIISGLFVSARRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 1053  
 QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTTQMTNTNDDLVGMQA 300  
 DB 1054 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTTQMTNTNDDLVGMQA 1113  
 QY 301 PGARSMCTCGSSDLVLTTRHADVIPIVRRGDSRGSLLSPRVSYLKSGSGPILCPSS 360  
 DB 1114 PPARSLTPTCTGSSDLVLTTRHADVIPIVRRGDSRGSLLSPRVSYLKSGSGPILCPSS 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 1206

## RESULT 11

US-09-263-933-11  
 / Sequence 11, Application US/09263933

Patent No. 6280940  
 / GENERAL INFORMATION:  
 / APPLICANT: Potts, Karen E.  
 / APPLICANT: Jackson, Roberta L.  
 / APPLICANT: Patrick, Amy K.  
 / TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 / TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 / FILE REFERENCE: 0125-0005A

QY 1 MAASCGAVFIQGLALTLSPYKVLARLWMLQYLITRVEAHQWIPPLNVRGSDAI 60  
 DB 93 MAASCGAVFVGLVLTLSPIYKFLARLWMLQYFTTRAEHLQWIPPLNVRGSDAI 152  
 QY 61 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAGGHY 120  
 DB 153 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAGGHY 212  
 QY 121 VQAMFKLAALTGYVVDHLTPLODMAHAGLDLAVAVEPVFSDETKIITWGADTAAC 180  
 DB 213 VQAMFKLAALTGYVVDHLTPLODMAHAGLDLAVAVEPVFSDETKIITWGADTAAC 272  
 QY 181 GDIISGLFVSARRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 240  
 DB 273 GDIISGLFVSARRGREILGPDNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDN 332  
 QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTTQMTNTNDDLVGMQA 300  
 DB 333 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTTQMTNTNDDLVGMQA 392  
 QY 301 PGARSMCTCGSSDLVLTTRHADVIPIVRRGDSRGSLLSPRVSYLKSGSGPILCPSS 360  
 DB 393 PPARSLTPTCTGSSDLVLTTRHADVIPIVRRGDSRGSLLSPRVSYLKSGSGPILCPSS 452  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 393  
 DB 453 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 485

## US-09-263-933-11

Query Match 94.6%; Score 1943; DB 3; Length 1692;  
 Best Local Similarity 92.6%; Pred. No. 3.8e-186;  
 Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

RESULT 12  
 US-09-919-901-11  
 / Sequence 11, Application US/09919901  
 / Patent No. 6599738  
 / GENERAL INFORMATION:  
 / APPLICANT: Potts, Karen E.  
 / APPLICANT: Jackson, Roberta L.  
 / APPLICANT: Patrick, Amy K.  
 / TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 / TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 / FILE REFERENCE: 0125-0005A  
 / CURRENT APPLICATION NUMBER: US/09/919,901  
 / CURRENT FILING DATE: 2001-08-02  
 / PRIOR APPLICATION NUMBER: 09/263,933  
 / PRIOR FILING DATE: 1999-02-08  
 / PRIOR APPLICATION NUMBER: 09/129,611  
 / PRIOR FILING DATE: 1998-08-05  
 / NUMBER OF SEQ ID NOS: 33  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 11  
 / LENGTH: 1692  
 / TYPE: PRT  
 / ORGANISM: Artificial Sequence  
 / FEATURE:  
 / OTHER INFORMATION: :  
 US-09-919-901-11

Query Match 94.6%; Score 1943; DB 4; Length 1692;

Best Local Similarity 92.6%; Pred. No. 3,6e-186;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKYLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAI 60  
DB 93 MAASCGAVFGLVLTLSPYKFLRLIMLQYFTTRABAHLMWIPPLNVRGGRDAI 152  
QY 61 ILTCAVHPELIFDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
DB 153 ILMCAVHPELIFDITKLLAIFGLPMLVQAGITRVPYFRAQGLIRACMLVRKAAGHY 212  
QY 121 VQMAFMKLAGLTYIYNHLTPLRDVAHAGLRDLAVAVEPVSDEMTKIIITGADTAAC 180  
DB 213 VQMAFMKLAGLTYIYNHLTPLRDVAHAGLRDLAVAVEPVSDEMTKIIITGADTAAC 272  
QY 181 GDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLIGCIIITSLTRDKN 240  
DB 273 GDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLIGCIIITSLTRDKN 332  
QY 241 QVEGEVQVSTATOSFLATCNGVCMTVFHGAAGSKTLAGKGPITQYTNVDDLVGMOA 300  
DB 333 QVEGEVQVSTATOSFLATCNGVCMTVFHGAAGSKTLAGKGPITQYTNVDDLVGMOA 392  
QY 301 PPGARSMTPCTCGSSDLYLTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCP 360  
DB 393 PPGARSMTPCTCGSSDLYLTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCP 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 485

RESULT 13  
US-09-263-933-9  
Sequence 9, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potte, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-9

Query Match 94.6%; Score 1943; DB 3; Length 2307;  
Best Local Similarity 92.6%; Pred. No. 6e-186;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKYLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAI 60  
DB 185 MAASCGAVFGLVLTLSPYKFLRLIMLQYFTTRABAHLMWIPPLNVRGGRDAI 244  
QY 61 ILTCAVHPELIFDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
DB 245 ILMCAVHPELIFDITKLLAIFGLPMLVQAGITRVPYFRAQGLIRACMLVRKAAGHY 304  
QY 121 VQMAFMKLAGLTYIYNHLTPLRDVAHAGLRDLAVAVEPVSDEMTKIIITGADTAAC 180  
DB 305 VQMAFMKLAGLTYIYNHLTPLRDVAHAGLRDLAVAVEPVSDEMTKIIITGADTAAC 364  
QY 181 GDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLIGCIIITSLTRDKN 240

DB 365 GDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLIGCIIITSLTRDKN 424  
QY 241 QVEGEVQVSTATOSFLATCNGVCMTVFHGAAGSKTLAGKGPITQYTNVDDLVGMOA 300  
DB 425 QVEGEVQVSTATOSFLATCNGVCMTVFHGAAGSKTLAGKGPITQYTNVDDLVGMOA 484  
QY 301 PPGARSMTPCTCGSSDLYLTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCP 360  
DB 485 PPGARSMTPCTCGSSDLYLTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCP 544  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 577

RESULT 14  
US-09-919-901-9  
Sequence 9, Application US/09919901  
Patent No. 6539738  
GENERAL INFORMATION:  
APPLICANT: Potte, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE: OTHER INFORMATION:  
US-09-919-901-9

Query Match 94.6%; Score 1943; DB 4; Length 2307;  
Best Local Similarity 92.6%; Pred. No. 6e-186;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKYLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAI 60  
DB 185 MAASCGAVFGLVLTLSPYKFLRLIMLQYFTTRABAHLMWIPPLNVRGGRDAI 244  
QY 61 ILTCAVHPELIFDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
DB 245 ILMCAVHPELIFDITKLLAIFGLPMLVQAGITRVPYFRAQGLIRACMLVRKAAGHY 304  
QY 121 VQMAFMKLAGLTYIYNHLTPLRDVAHAGLRDLAVAVEPVSDEMTKIIITGADTAAC 180  
DB 305 VQMAFMKLAGLTYIYNHLTPLRDVAHAGLRDLAVAVEPVSDEMTKIIITGADTAAC 364  
QY 181 GDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLIGCIIITSLTRDKN 240  
DB 365 GDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLIGCIIITSLTRDKN 424  
QY 241 QVEGEVQVSTATOSFLATCNGVCMTVFHGAAGSKTLAGKGPITQYTNVDDLVGMOA 300  
DB 425 QVEGEVQVSTATOSFLATCNGVCMTVFHGAAGSKTLAGKGPITQYTNVDDLVGMOA 484  
QY 301 PPGARSMTPCTCGSSDLYLTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCP 360  
DB 485 PPGARSMTPCTCGSSDLYLTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCP 544  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 577

RESULT 15  
US-09-263-933-18  
; Sequence 18, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/263,933  
; CURRENT FILING DATE: 1999-03-08  
; EARLIER APPLICATION NUMBER: 09/129,611  
; EARLIER FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; US-09-263-933-18

Query Match 94.2%; Score 1934; DB 3; Length 1692;  
Best Local Similarity 92.4%; Pred. No. 3, 1e-185;  
Matches 363; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

QY	1	MAASCGAVFIGLALTLSPFYVLLARLIWLOYLITREAHLOVWIPPLAVRGGRDAI	60
DB	93	MAASCGAVVGVGLVLTSPYKVFLLARLIWLOYLITREAHLOVWIPPLAVRGGRDAI	152
QY	61	ILTCVHPPELIPDITKLALIFGPIMLVQAGITKVFYFVRAOGLIRACMLYKKAAGHY	120
DB	153	ILNCAVHPPELIPDITKLALIFGPIMLVQAGITKVFYFVRAOGLIRACMLYKKAAGHY	212
QY	121	VQNAFMKLAALGTYYVDHLTPLQDMAHAGLRDLAVAVEPVFSDMEVKITTWGADTAAC	180
DB	213	VQNAFMKLAALGTYYVDHLTPLQDMAHAGLRDLAVAVEPVFSDMEVKITTWGADTAAC	272
QY	181	GDITSGIPVARGREILIGPADNFEQGWRLAPITAYSOOTRGILGCIITSLTGRDKY	240
DB	273	GDITSGIPVARGREILIGPADNFEQGWRLAPITAYSOOTRGILGCIITSLTGRDKY	332
QY	241	QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDQDLVGMQA	300
DB	333	QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDQDLVGMQA	392
QY	301	PGARSKVTPCTCGSSDLYLTRHADVTVRRRSDSGSLSPRVSYLTKSSGGPILCPG	360
DB	393	PGARSKVTPCTCGSSDLYLTRHADVTVRRRSDSGSLSPRVSYLTKSSGGPILCPG	452
QY	361	GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR	393
DB	453	GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR	485

Search completed: May 6, 2004, 09:39:02  
Job time: 16.4246 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 40.2025 Seconds  
(without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MAASCGAVFGLALTLPSP.....RGVAKVDFPESMETTR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2053	100.0	393	13	US-10-017-736-11
2	2053	100.0	393	13	US-10-650-585-11
3	2053	100.0	409	16	US-10-017-736-2
4	2053	100.0	409	16	US-10-650-585-2
5	1987	96.8	380	13	US-10-017-736-12
6	1987	96.8	380	13	US-10-650-585-12
7	1951	95.0	2201	13	US-10-029-807-3
8	1951	95.0	2201	13	US-10-309-561-3
9	1951	95.0	3010	12	US-10-467-000-1
10	1946	94.8	1692	14	US-09-919-901-4
11	1946	94.8	1692	14	US-10-191-966-4
12	1946	94.8	2307	10	US-09-919-901-2
13	1946	94.8	2307	14	US-10-191-966-2
14	1943	94.6	1692	10	US-09-919-901-1
15	1943	94.6	1692	14	US-10-191-966-1

ALIGNMENTS

16	1943	94.6	2307	10	US-09-919-901-9	Sequence 9, Appl1
17	1943	94.6	2307	14	US-10-191-966-9	Sequence 9, Appl1
18	1934	94.2	1692	10	US-09-919-901-18	Sequence 18, Appl1
19	1934	94.2	1692	14	US-10-191-966-18	Sequence 18, Appl1
20	1934	94.2	2307	10	US-09-919-901-16	Sequence 16, Appl1
21	1934	94.2	2307	14	US-10-191-966-16	Sequence 16, Appl1
22	1888	92.0	2201	13	US-10-085-476-2	Sequence 2, Appl1
23	1842	89.7	352	13	US-10-017-736-13	Sequence 13, Appl1
24	1842	89.7	352	16	US-10-650-585-13	Sequence 13, Appl1
25	1778	86.6	341	13	US-10-017-736-14	Sequence 14, Appl1
26	1778	86.6	341	16	US-10-650-585-14	Sequence 14, Appl1
27	1772	86.3	2985	14	US-10-259-275-10	Sequence 40, Appl1
28	1766	86.0	3011	9	US-09-742-659-4	Sequence 4, Appl1
29	1766	86.0	3011	10	US-09-891-894-3	Sequence 3, Appl1
30	1766	86.0	3011	15	US-10-184-150-3	Sequence 3, Appl1
31	1766	86.0	3011	14	US-10-328-997-3	Sequence 3, Appl1
32	1766	86.0	3012	9	US-09-238-076-2	Sequence 2, Appl1
33	1766	86.0	3012	10	US-09-995-937-2	Sequence 2, Appl1
34	1766	86.0	3012	10	US-09-917-563-2	Sequence 2, Appl1
35	1764	85.9	3011	9	US-09-916-359-2	Sequence 2, Appl1
36	1764	85.9	3011	12	US-10-296-734-406	Sequence 406, App
37	1762	85.8	3011	9	US-09-238-076-20	Sequence 20, Appl1
38	1762	85.8	3011	10	US-09-995-937-20	Sequence 20, Appl1
39	1762	85.8	3011	10	US-09-917-563-20	Sequence 20, Appl1
40	1759	85.7	2894	9	US-09-941-611-23	Sequence 23, Appl1
41	1759	85.7	2894	14	US-10-044-995-23	Sequence 23, Appl1
42	1759	85.7	3011	9	US-09-952-572-8	Sequence 9, Appl1
43	1759	85.7	3011	9	US-09-747-419-20	Sequence 9, Appl1
44	1759	85.7	3011	12	US-10-189-359-14	Sequence 14, Appl1
45	1759	85.7	3011	14	US-10-259-275-20	Sequence 20, Appl1

RESULT 1  
US-10-017-736-11  
Sequence 11, Application US/10017736  
Publication No. US20020192640A1  
GENERAL INFORMATION:  
APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
FILE REFERENCE: 13/082  
CURRENT APPLICATION NUMBER: US/10/017,736  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: 60/256,031  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11  
LENGTH: 393  
TYPE: PRT  
ORGANISM: HCV  
US-10-017-736-11

Query Match 100.0%; Score 2053; DB 13; Length 393;  
Best Local Similarity 100.0%; Pred. No. 2.2e-197;  
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAASCGAVFGLALTLPSPYKVLARLIMWLQYLRVREAHQVMIPLNVRGGRDAI	60
DB	1	MAASCGAVFGLALTLPSPYKVLARLIMWLQYLRVREAHQVMIPLNVRGGRDAI	60
QY	61	ILTCVHPELFDITKLLAIFGPMVLAQITKPYFVRAGLIRACMVRKAAGHY	120
DB	61	ILTCVHPELFDITKLLAIFGPMVLAQITKPYFVRAGLIRACMVRKAAGHY	120
QY	121	VQMAFKALALGTYYDHLTFLQWNAHGLFDLVANVPEVFSMEVYLIIRWADTAC	180
DB	121	VQMAFKALALGTYYDHLTFLQWNAHGLFDLVANVPEVFSMEVYLIIRWADTAC	180
QY	181	GDIIISGLPVSARGREIILGPAIDFEGQGRILAPITAYSOQTRGLGCIITSLGRDKI	240

Db 181 GDIISGLPVBARREIILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 240  
Qy 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Db 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Qy 301 PPGARSWTPTCGSSDLVYTRHADVI PVRRRGSGLSPRPVSYLKSSGGPILCPs 360  
Db 301 PPGARSWTPTCGSSDLVYTRHADVI PVRRRGSGLSPRPVSYLKSSGGPILCPs 360  
Qy 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

## RESULT 2

US-10-650-585-11  
; Sequence 11, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; PRIORITY FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 393  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-650-585-11

Query Match 100.0%; Score 2053; DB 16; Length 393;  
Best Local Similarity 100.0%; Pred. No. 2, 2e-197;

Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHQVWIPPLNVRGGDAI 60  
Db 1 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHQVWIPPLNVRGGDAI 60  
Qy 61 ILITCAVHPELIFDITKLLAIFGPLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
Db 61 ILITCAVHPELIFDITKLLAIFGPLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
Qy 121 VQMAFMKLAALTGYYVDHLTPLODWAHAGLRDLA VAEVPIFSDEMEVKIITWGADTAAC 180  
Db 121 VQMAFMKLAALTGYYVDHLTPLODWAHAGLRDLA VAEVPIFSDEMEVKIITWGADTAAC 180  
Qy 181 GDIISGLPVBARREIILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 240  
Db 181 GDIISGLPVBARREIILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 240  
Qy 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Db 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Qy 301 PPGARSWTPTCGSSDLVYTRHADVI PVRRRGSGLSPRPVSYLKSSGGPILCPs 360  
Db 301 PPGARSWTPTCGSSDLVYTRHADVI PVRRRGSGLSPRPVSYLKSSGGPILCPs 360  
Qy 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

## RESULT 3

US-10-017-736-2  
; Sequence 2, Application US/10017736

; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIORITY FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-017-736-2

Query Match 100.0%; Score 2053; DB 13; Length 409;  
Best Local Similarity 100.0%; Pred. No. 2, 3e-197;  
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHQVWIPPLNVRGGDAI 60  
Db 5 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHQVWIPPLNVRGGDAI 64  
Qy 61 ILITCAVHPELIFDITKLLAIFGPLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
Db 65 ILITCAVHPELIFDITKLLAIFGPLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 124  
Qy 121 VQMAFMKLAALTGYYVDHLTPLODWAHAGLRDLA VAEVPIFSDEMEVKIITWGADTAAC 180  
Db 125 VQMAFMKLAALTGYYVDHLTPLODWAHAGLRDLA VAEVPIFSDEMEVKIITWGADTAAC 184  
Qy 181 GDIISGLPVBARREIILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 240  
Db 185 GDIISGLPVBARREIILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 244  
Qy 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Db 245 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 304  
Qy 301 PPGARSWTPTCGSSDLVYTRHADVI PVRRRGSGLSPRPVSYLKSSGGPILCPs 360  
Db 305 PPGARSWTPTCGSSDLVYTRHADVI PVRRRGSGLSPRPVSYLKSSGGPILCPs 364  
Qy 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397

## RESULT 4

US-10-650-585-2  
; Sequence 2, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; PRIORITY FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-650-585-2

Query Match 100.0%; Score 2053; DB 16; Length 409;

Best Local Similarity 100.0%; Pred. No. 2,3e-197;  
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAASCGAVFIQALITLSPYKVLARLIWLOYLITVEAHLQWIPPLNVRGGRDAI 60  
DB 5 MAASCGAVFIQALITLSPYKVLARLIWLOYLITVEAHLQWIPPLNVRGGRDAI 64  
QY 61 ILTTCVHPELLFDITKLALIFGRLMVLQAGITKVPYFRAOGLIRACMLVRKAAGHY 120  
DB 65 ILTTCVHPELLFDITKLALIFGRLMVLQAGITKVPYFRAOGLIRACMLVRKAAGHY 124  
QY 121 VQVAFMKALITGYVDHLTPLOQWMAHGLRDLAVAVEPIFSDMEVKIITWGADTAAC 180  
DB 125 VQVAFMKALITGYVDHLTPLOQWMAHGLRDLAVAVEPIFSDMEVKIITWGADTAAC 184  
QY 181 GDIISGLPVSARRGRELILGPADNFGQWRLLAPITAYSQOTRGLGCIITSLGRDKN 240  
DB 185 GDIISGLPVSARRGRELILGPADNFGQWRLLAPITAYSQOTRGLGCIITSLGRDKN 244  
QY 241 OVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKGPITQMTNVODLVGMQA 300  
DB 245 OVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKGPITQMTNVODLVGMQA 304  
QY 301 PGARSMTPCTCGSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCP 360  
DB 305 PGARSMTPCTCGSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCP 364  
QY 361 GHAVGFRAVCTRGVAKVDFIPVESMETMR 393  
DB 365 GHAVGFRAVCTRGVAKVDFIPVESMETMR 397

RESULT 5  
US-10-017-736-12

; Sequence 12, Application US/10017736  
; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-017-736-12

Query Match 96.8%; Score 1987; DB 13; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9e-191;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ALLTSPYKVLARLIWLOYLITVEAHLQWIPPLNVRGGRDAIILTCVHPELLF 73  
DB 1 ALLTSPYKVLARLIWLOYLITVEAHLQWIPPLNVRGGRDAIILTCVHPELLF 60  
QY 74 DITKLALIFGRLMVLQAGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFMKLAALTG 133  
DB 61 DITKLALIFGRLMVLQAGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFMKLAALTG 120  
QY 134 TYVDHLTPLOQWMAHGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVSARR 193  
DB 121 TYVDHLTPLOQWMAHGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVSARR 180  
QY 194 GREILGPADNFGQWRLLAPITAYSQOTRGLGCIITSLGRDKNQVEGEVQVSTAT 253  
DB 181 GREILGPADNFGQWRLLAPITAYSQOTRGLGCIITSLGRDKNQVEGEVQVSTAT 240  
QY 254 OSFLATCVNGVCMVTFHAGSKTLAPKGPITQMTNVODLVGMQAPPGARSMTPCTCG 313

DB 241 OSFLATCVNGVCMVTFHAGSKTLAPKGPITQMTNVODLVGMQAPPGARSMTPCTCG 300

QY 314 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGHAAGVIFRAAVCT 373  
DB 301 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGHAAGVIFRAAVCT 360  
QY 374 RGVAKAVDFIPVESMETMR 393  
DB 361 RGVAKAVDFIPVESMETMR 380

RESULT 6  
US-10-650-585-12

; Sequence 12, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR FILING DATE: US/10/017,736A  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-650-585-12

Query Match 96.8%; Score 1987; DB 16; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9e-191;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ALLTSPYKVLARLIWLOYLITVEAHLQWIPPLNVRGGRDAIILTCVHPELLF 73  
DB 1 ALLTSPYKVLARLIWLOYLITVEAHLQWIPPLNVRGGRDAIILTCVHPELLF 60  
QY 74 DITKLALIFGRLMVLQAGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFMKLAALTG 133  
DB 61 DITKLALIFGRLMVLQAGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFMKLAALTG 120  
QY 134 TYVDHLTPLOQWMAHGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVSARR 193  
DB 121 TYVDHLTPLOQWMAHGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVSARR 180  
QY 194 GREILGPADNFGQWRLLAPITAYSQOTRGLGCIITSLGRDKNQVEGEVQVSTAT 253  
DB 181 GREILGPADNFGQWRLLAPITAYSQOTRGLGCIITSLGRDKNQVEGEVQVSTAT 240  
QY 254 OSFLATCVNGVCMVTFHAGSKTLAPKGPITQMTNVODLVGMQAPPGARSMTPCTCG 313  
DB 241 OSFLATCVNGVCMVTFHAGSKTLAPKGPITQMTNVODLVGMQAPPGARSMTPCTCG 300  
QY 314 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGHAAGVIFRAAVCT 373  
DB 301 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGHAAGVIFRAAVCT 360  
QY 374 RGVAKAVDFIPVESMETMR 393  
DB 361 RGVAKAVDFIPVESMETMR 380

RESULT 7  
US-10-029-907-3

; Sequence 3, Application US/10029907  
; Publication No. US20020142350A1  
; GENERAL INFORMATION:  
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.

```

; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
; US-10-029-907-3

```

```

Query Match          95.0%; Score 1951; DB 13; Length 2201;
Best Local Similarity 93.1%; Pred. No. 4,4e-186;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 1 MAASCGAVFTGLAL:TLSPYKVLARLIMLOYLITREVAHLQWIPPLNVRGRDAI 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 5 MAASCGAVFVGLILTLSPHYKFLARLIMLOYLITREVAHLQWIPPLNVRGRDAV 64
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 ILTCAVHPELIFDITKLLAIFQPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 65 ILTCAHPELIFITKLLAIFPLMVLQAGITKVPYFVRAHGLIRACMLVRKAGHY 124
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 VQNAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEVPSDMEVKIITWGADTAAC 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 125 VQNAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEVPSDMEVKIITWGADTAAC 184
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 GDIISGLPVSARRGREIHLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLGRDN 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 185 GDIISGLPVSARRGREIHLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLGRDN 244
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAPKXGITQWYNVDDLVGMQA 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 245 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAPKXGITQWYNVDDLVGMQA 304
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 PPGARSMPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 305 PPGARSLTPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 364
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 397
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 8
US-10-309-561-3
; Sequence 3, Application US/10309561
; Publication No. US20030148348A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US/10/029,907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3

```

```

; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
; US-10-309-561-3

```

```

Query Match          95.0%; Score 1951; DB 14; Length 2201;
Best Local Similarity 93.1%; Pred. No. 4,4e-186;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 1 MAASCGAVFTGLAL:TLSPYKVLARLIMLOYLITREVAHLQWIPPLNVRGRDAI 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 5 MAASCGAVFVGLILTLSPHYKFLARLIMLOYLITREVAHLQWIPPLNVRGRDAV 64
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 ILTCAVHPELIFDITKLLAIFQPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 65 ILTCAHPELIFITKLLAIFPLMVLQAGITKVPYFVRAHGLIRACMLVRKAGHY 124
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 VQNAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEVPSDMEVKIITWGADTAAC 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 125 VQNAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEVPSDMEVKIITWGADTAAC 184
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 GDIISGLPVSARRGREIHLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLGRDN 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 185 GDIISGLPVSARRGREIHLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLGRDN 244
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAPKXGITQWYNVDDLVGMQA 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 245 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAPKXGITQWYNVDDLVGMQA 304
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 PPGARSMPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 305 PPGARSLTPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 364
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 397
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 9
US-10-467-000-1
; Sequence 1, Application US/10467000
; Publication No. US20040067486A1
; GENERAL INFORMATION:
; APPLICANT: De Francesco, Raffaele
; APPLICANT: Migliaccio, Giovanni
; APPLICANT: Paolesse, Giacomo
; TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
; FILE REFERENCE: ITR0003P
; CURRENT APPLICATION NUMBER: US/10/467,000
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: PCT/EP02/00526
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/263,479
; PRIOR FILING DATE: 2001-01-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Con 1 HCV isolate nucleic acid
; US-10-467-000-1

```

```

Query Match          95.0%; Score 1951; DB 12; Length 3010;
Best Local Similarity 93.1%; Pred. No. 6,9e-186;

```

Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```

QY 1 MAASCGAVFGLALITLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60
DB 814 MAASCGAVFGLITLSPHYKLFLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 873
QY 61 ILITCAVHPELITDKLLAIFGPIMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 120
DB 874 ILITCAVHPELITDKLLAIFGPIMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 933
QY 121 VQAFMKLAALTGTYYVDHLTPQDMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 180
DB 934 VQAFMKLAALTGTYYVDHLTPQDMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 993
QY 181 GDIISGLPVASARGREIILGPADNREGQWRLAPITAYSQOTRGLGCIITSLTGRDN 240
DB 994 GDIISGLPVASARGREIILGPADNREGQWRLAPITAYSQOTRGLGCIITSLTGRDN 1053
QY 241 QVEGEVQVSTATQSFATCVNGCMTVPHGASKTLAEPKGPITOMYTNVDODLVGMOA 300
DB 1054 QVEGEVQVSTATQSFATCVNGCMTVPHGASKTLAEPKGPITOMYTNVDODLVGMOA 1113
QY 301 PPGASMTPTCTGSSDLYLVTRHADVIIVRRRDSRGLSPRPVSYLKSGSGGPLLCP 360
DB 1114 PPGASMTPTCTGSSDLYLVTRHADVIIVRRRDSRGLSPRPVSYLKSGSGGPLLCP 1173
QY 361 GHAVGIFRAAVCTRGVAKAVDFIVPESMETMR 393
DB 1174 GHAVGIFRAAVCTRGVAKAVDFIVPESMETMR 1206

```

## RESULT 10

```

US-09-919-901-4
; Sequence 4, Application US/09919901
; Publication No. US2003082518A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-4

```

Query Match 94.8%; Score 1946; DB 10; Length 1692;  
 Best Local Similarity 92.9%; Pred. No. 9,7e-186;  
 Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

```

QY 1 MAASCGAVFGLALITLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60
DB 93 MAASCGAVFGLVLLTSLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 152
QY 61 ILITCAVHPELITDKLLAIFGPIMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 120
DB 153 ILITCAVHPELITDKLLAIFGPIMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 212
QY 121 VQAFMKLAALTGTYYVDHLTPQDMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 180
DB 213 VQAFMKLAALTGTYYVDHLTPQDMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 272

```

```

QY 181 GDIISGLPVASARGREIILGPADNREGQWRLAPITAYSQOTRGLGCIITSLTGRDN 240
DB 273 GDIISGLPVASARGREIILGPADNREGQWRLAPITAYSQOTRGLGCIITSLTGRDN 332
QY 241 QVEGEVQVSTATQSFATCVNGCMTVPHGASKTLAEPKGPITOMYTNVDODLVGMOA 300
DB 333 QVEGEVQVSTATQSFATCVNGCMTVPHGASKTLAEPKGPITOMYTNVDODLVGMOA 392
QY 301 PPGASMTPTCTGSSDLYLVTRHADVIIVRRRDSRGLSPRPVSYLKSGSGGPLLCP 360
DB 393 PPGASMTPTCTGSSDLYLVTRHADVIIVRRRDSRGLSPRPVSYLKSGSGGPLLCP 452
QY 361 GHAVGIFRAAVCTRGVAKAVDFIVPESMETMR 393
DB 453 GHAVGIFRAAVCTRGVAKAVDFIVPESMETMR 485

```

## RESULT 11

```

US-10-191-966-4
; Sequence 4, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-10-191-966-4

```

Query Match 94.8%; Score 1946; DB 14; Length 1692;  
 Best Local Similarity 92.9%; Pred. No. 9,7e-186;  
 Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

```

QY 1 MAASCGAVFGLALITLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60
DB 93 MAASCGAVFGLVLLTSLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 152
QY 61 ILITCAVHPELITDKLLAIFGPIMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 120
DB 153 ILITCAVHPELITDKLLAIFGPIMLVQAGITKVPYFRAQGLIRACMLVRKAAGHY 212
QY 121 VQAFMKLAALTGTYYVDHLTPQDMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 180
DB 213 VQAFMKLAALTGTYYVDHLTPQDMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 272
QY 241 QVEGEVQVSTATQSFATCVNGCMTVPHGASKTLAEPKGPITOMYTNVDODLVGMOA 300
DB 333 QVEGEVQVSTATQSFATCVNGCMTVPHGASKTLAEPKGPITOMYTNVDODLVGMOA 392
QY 301 PPGASMTPTCTGSSDLYLVTRHADVIIVRRRDSRGLSPRPVSYLKSGSGGPLLCP 360
DB 393 PPGASMTPTCTGSSDLYLVTRHADVIIVRRRDSRGLSPRPVSYLKSGSGGPLLCP 452

```

QY 361 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
DB 453 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 485

## RESULT 12

US-09-919-901-2  
Sequence 2, Application US/09919901  
Publication No. US2003082518A1  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PRF  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-2

Query Match 94.8%; Score 1946; DB 10; Length 2307;  
Best Local Similarity 92.9%; Pred. No. 1.5e-185;

Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60  
DB 185 MAASCGAVFVGLVLLTSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 244  
QY 61 ILTCAVHPELIPITILLAIKPLAVLQAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
DB 245 ILMCAVHPELIPITILLAIKPLAVLQAGITRVPYFRAQGLIRACMLVRKAGGHY 304  
QY 121 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKIITMGADTAAC 180  
DB 305 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKIITMGADTAAC 364  
QY 181 GDIISGLPVASARRGRELILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 240  
DB 365 GDIISGLPVASARRGRELILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 424  
QY 241 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGKPIITOMYTNVQDVLVQMA 300  
DB 425 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGKPIITOMYTNVQDVLVQMA 484  
QY 301 PPGARSMTPCTCGSSDLYLVTRADVIYVRRRGDSRGLSPRPVSYLKSSGGPLLCPS 360  
DB 485 PPGARSMTPCTCGSSDLYLVTRADVIYVRRRGDSRGLSPRPVSYLKSSGGPLLCPS 544  
QY 361 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
DB 545 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 577

## RESULT 13

US-10-191-966-2  
Sequence 2, Application US/10191966  
Publication No. US20030175692A1  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.

;; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
;; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
;; FILE REFERENCE: 0125-0005A  
;; CURRENT APPLICATION NUMBER: US/10/191,966  
;; CURRENT FILING DATE: 2002-07-10  
;; PRIOR APPLICATION NUMBER: US/09/263,933  
;; PRIOR FILING DATE: 1999-03-08  
;; PRIOR APPLICATION NUMBER: 09/129,611  
;; PRIOR FILING DATE: 1998-08-05  
;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 2  
;; LENGTH: 2307  
;; TYPE: PRF  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: :  
US-10-191-966-2

Query Match 94.8%; Score 1946; DB 14; Length 2307;  
Best Local Similarity 92.9%; Pred. No. 1.5e-185;  
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60  
DB 185 MAASCGAVFVGLVLLTSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 244  
QY 61 ILTCAVHPELIPITILLAIKPLAVLQAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
DB 245 ILMCAVHPELIPITILLAIKPLAVLQAGITRVPYFRAQGLIRACMLVRKAGGHY 304  
QY 121 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKIITMGADTAAC 180  
DB 305 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKIITMGADTAAC 364  
QY 181 GDIISGLPVASARRGRELILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 240  
DB 365 GDIISGLPVASARRGRELILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 424  
QY 241 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGKPIITOMYTNVQDVLVQMA 300  
DB 425 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGKPIITOMYTNVQDVLVQMA 484  
QY 301 PPGARSMTPCTCGSSDLYLVTRADVIYVRRRGDSRGLSPRPVSYLKSSGGPLLCPS 360  
DB 485 PPGARSMTPCTCGSSDLYLVTRADVIYVRRRGDSRGLSPRPVSYLKSSGGPLLCPS 544  
QY 361 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
DB 545 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 577

## RESULT 14

US-09-919-901-11  
Sequence 11, Application US/09919901  
Publication No. US2003082518A1  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-11

Query Match 94.6%; Score 1943; DB 10; Length 1692;  
Best Local Similarity 92.6%; Pred. No. 1.9e-185;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGGAVFGIALLTSPYKVLARLIMWLQYITVEAHLQVMIPELVNRRGRDAI 60  
DB 93 MAASCGGAVFGIALLTSPYKVLARLIMWLQYITVEAHLQVMIPELVNRRGRDAI 152  
QY 61 ILTCAVHPELFDITKLLAIIFGRLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHY 120  
DB 153 ILMCAVHPELFDITKLLAIIFGRLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHY 212  
QY 121 VQMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEPVFSDMEVKIITWGADTAAC 180  
DB 213 VQMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEPVFSDMETKIIITWGADTAAC 272  
QY 181 GDIISGLPVSAARRGRELILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDXN 240  
DB 273 GDIISGLPVSAARRGRELILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDXN 332  
QY 241 QVEGEVQVSTATOSFLATCVNGVCTVPHGAGSKTLAGPKGPIITQMTYNVDODLVGMQA 300  
DB 333 QVEGEVQVSTATOSFLATCVNGVCTVPHGAGSKTLAGPKGPIITQMTYNVDODLVGMQA 392  
QY 301 PPGARSMTPCTCGSSDLYLVTRHADVIIVRRRGRSRLSPRVSYLKSGSGGFLCPS 360  
DB 393 PPGARSLTPCTCGSSDLYLVTRHADVIIVRRRGRSRLSPRVSYLKSGAGGFLCPS 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 485

RESULT 15  
US-10-191-966-11  
; Sequence 11, Application US/10191966  
; Publication No. US20030175692A1  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/10/191,966  
; CURRENT FILING DATE: 2002-07-10  
; PRIOR APPLICATION NUMBER: US/09/263,933  
; PRIOR FILING DATE: 1999-03-08  
; PRIOR APPLICATION NUMBER: 09/129,611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 11.  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: :  
US-10-191-966-11

Query Match 94.6%; Score 1943; DB 14; Length 1692;  
Best Local Similarity 92.6%; Pred. No. 1.9e-185;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGGAVFGIALLTSPYKVLARLIMWLQYITVEAHLQVMIPELVNRRGRDAI 60  
DB 93 MAASCGGAVFGIALLTSPYKVLARLIMWLQYITVEAHLQVMIPELVNRRGRDAI 152

QY 61 ILTCAVHPELFDITKLLAIIFGRLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHY 120  
DB 153 ILMCAVHPELFDITKLLAIIFGRLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHY 212  
QY 121 VQMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEPVFSDMEVKIITWGADTAAC 180  
DB 213 VQMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAVEPVFSDMETKIIITWGADTAAC 272  
QY 181 GDIISGLPVSAARRGRELILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDXN 240  
DB 273 GDIISGLPVSAARRGRELILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDXN 332  
QY 241 QVEGEVQVSTATOSFLATCVNGVCTVPHGAGSKTLAGPKGPIITQMTYNVDODLVGMQA 300  
DB 333 QVEGEVQVSTATOSFLATCVNGVCTVPHGAGSKTLAGPKGPIITQMTYNVDODLVGMQA 392  
QY 301 PPGARSMTPCTCGSSDLYLVTRHADVIIVRRRGRSRLSPRVSYLKSGSGGFLCPS 360  
DB 393 PPGARSLTPCTCGSSDLYLVTRHADVIIVRRRGRSRLSPRVSYLKSGAGGFLCPS 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 485

Search completed: May 6, 2004, 09:43:19  
Job time : 41.2025 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 51.7244 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-12  
Sequence: 1 ALLTSPYKYLARLIMWL.....RGVAKAVDFPVEHMETTR 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29qand4:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1987	100.0	380	5	ABG32185 HCV prote
2	1987	100.0	393	5	ABG32184 HCV prote
3	1987	100.0	409	5	ABG32181 HCV prote
4	1902	95.7	3010	2	AAR8694 Partial H
5	1897	95.5	3010	2	AAR8682 HCV prote
6	1896	95.4	3010	2	AAR8684 HCV prote
7	1888	95.0	768	2	AAR40223 Recombina
8	1887	95.0	2201	5	ABG30601 Hepatitis
9	1887	95.0	2201	5	ABG30591 Hepatitis
10	1887	95.0	2201	5	ABG30600 Hepatitis
11	1887	95.0	2201	5	ABG30581 Hepatitis
12	1887	95.0	2201	5	ABG30593 Hepatitis
13	1887	95.0	2201	5	ABG30582 Hepatitis
14	1887	95.0	2201	5	ABG30580 Hepatitis
15	1887	95.0	2201	5	ABG30587 Hepatitis
16	1887	95.0	2201	5	ABG30599 Hepatitis
17	1887	95.0	2201	5	ABG30594 Hepatitis
18	1887	95.0	2201	5	ABG30598 Hepatitis
19	1887	95.0	2201	5	ABG30595 Hepatitis
20	1887	95.0	3010	5	ABG32458 Hepatitis
21	1887	95.0	3010	5	ABG32459 Hepatitis
22	1887	95.0	3010	5	ABG32451 Hepatitis
23	1887	95.0	3010	5	ABG32455 Hepatitis
24	1887	95.0	3010	5	ABG32457 Hepatitis
25	1887	95.0	3010	5	ABG32460 Hepatitis

26	1887	95.0	3010	5	ABG32461 Hepatitis
27	1887	95.0	3010	5	ABG32454 Hepatitis
28	1887	95.0	3011	5	ABG32456 Hepatitis
29	1884	94.8	2201	5	ABG30586 Hepatitis
30	1884	94.8	2201	5	ABG30589 Hepatitis
31	1884	94.8	2201	5	ABG30583 Hepatitis
32	1884	94.8	2201	5	ABG30588 Hepatitis
33	1883	94.8	2201	5	ABG30590 Hepatitis
34	1881	94.7	2307	3	ABG32452 Hepatitis
35	1881	94.7	3010	5	ABG30584 Hepatitis
36	1880	94.6	2201	5	ABG30582 Hepatitis
37	1880	94.6	2201	5	ABG30602 Hepatitis
38	1880	94.6	3010	5	ABG32453 Hepatitis
39	1878	94.5	2307	3	AAV70065 Recombina
40	1878	94.5	3014	2	AAV54099 NANBHV E1
41	1876	94.5	2201	5	ABG30585 Hepatitis
42	1873	94.3	3014	2	AAR35207 Hepatitis
43	1872	94.2	3090	7	ADD67962 EMCV inte
44	1869	94.1	2307	3	AAV70066 Recombina
45	1869	94.1	3010	2	AAW98022 Infection

## ALIGNMENTS

RESULT 1  
ID ABG32185 standard; protein; 380 AA.  
XX  
AC ABG32185;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE HCV protease NS2/3 truncation mutant 827-1206.  
XX  
XX HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
KW chronic liver disease; cirrhosis; end-stage liver disease; virologic;  
KW hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; mutain.  
XX  
XX Hepatitis C virus.  
OS Synthetic.  
OS  
XX WO200248375-A2.  
XX  
XX 20-JUN-2002.  
XX  
XX 13-DEC-2001; 2001MO-CA001796.  
XX  
XX 15-DEC-2000; 2000US-0256031P.  
XX  
XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
XX Thibault D, Lamare D, Maurice R, Pilote L, Pause A;  
XX WPI, 2002-599511/64.  
XX  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.  
XX  
XX Claim 41; Page 60-61; 67bp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
XX CC residue residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
XX its truncation or a mutated sequence, where the protease is in a solution  
XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32184; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 827-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 380 AA;

Query Match 100.0%; Score 1987; DB 5; Length 380;

Best Local Similarity 100.0%; Pred. No. 1.5e-184; Indels 0; Gaps 0;

Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGRDAIILLTCVHPELIF 60  
 Db 1 ALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGRDAIILLTCVHPELIF 60  
 QY 61 DITKLLAIFGPIMLVQAGITKPYFVRAGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120  
 Db 61 DITKLLAIFGPIMLVQAGITKPYFVRAGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120  
 QY 121 TYYVDHLTFLQDWAHAGLNDLAVAVEPVFSDMEVKIITWGADTACGDIISGLPVSAR 180  
 Db 121 TYYVDHLTFLQDWAHAGLNDLAVAVEPVFSDMEVKIITWGADTACGDIISGLPVSAR 180  
 QY 121 TYYVDHLTFLQDWAHAGLNDLAVAVEPVFSDMEVKIITWGADTACGDIISGLPVSAR 180  
 Db 121 TYYVDHLTFLQDWAHAGLNDLAVAVEPVFSDMEVKIITWGADTACGDIISGLPVSAR 180  
 QY 181 GREILLGPADNFEQGQWRLIAPITAYSSQTRGLLGCIITSLTGRDNQYEGEVQVSTAT 240  
 Db 181 GREILLGPADNFEQGQWRLIAPITAYSSQTRGLLGCIITSLTGRDNQYEGEVQVSTAT 240  
 QY 241 QSEFLATCVAGVCTVTHGASGKTLAGPKGITOMTNNVDDLVGMQAPPGASMTPTCG 300  
 Db 241 QSEFLATCVAGVCTVTHGASGKTLAGPKGITOMTNNVDDLVGMQAPPGASMTPTCG 300  
 QY 301 SSDLVLTVRHADVIPIVRRGDSRGSILSPRVSYLKSSGGPDLCPGHAIVGIFRAAVCT 360  
 Db 301 SSDLVLTVRHADVIPIVRRGDSRGSILSPRVSYLKSSGGPDLCPGHAIVGIFRAAVCT 360  
 QY 361 RGVAKAVDPIPVESMETTR 380  
 Db 361 RGVAKAVDPIPVESMETTR 380

RESULT 2

ABG32184

ID ABG32184 standard; protein; 393 AA.

XX ABG32184;

DT 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 815-1206.

KM HCV, enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viricide;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.

XX Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200248375-A2.  
 XX  
 PD 20-JUN-2002.  
 XX  
 PF 13-DEC-2001; 2001WO-CA001796.  
 XX  
 PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-599511/64.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS  
 PS Claim 41; Page 59-60; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-  
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 residue amino acid 810 to 906, or having a minimal amino acid sequence  
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 NS2/3 protease. Also included are (1) a composition (C) comprising an  
 isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 its truncation or a mutated sequence, where the protease is in a solution  
 comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 appearing as ABG32184; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein)

SO Sequence 393 AA;

Query Match 100.0%; Score 1987; DB 5; Length 393;

Best Local Similarity 100.0%; Pred. No. 1.5e-184; Indels 0; Gaps 0;

Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGRDAIILLTCVHPELIF 60  
 Db 1 ALTLSPYKVLARLIMWLYITRVAHLQVWIPPLNVRGGRDAIILLTCVHPELIF 60  
 QY 61 DITKLLAIFGPIMLVQAGITKPYFVRAGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120  
 Db 61 DITKLLAIFGPIMLVQAGITKPYFVRAGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120  
 QY 74 DITKLLAIFGPIMLVQAGITKPYFVRAGLIRACMLVRKKAAGHYVQMAFMKLAALTG 133  
 Db 74 DITKLLAIFGPIMLVQAGITKPYFVRAGLIRACMLVRKKAAGHYVQMAFMKLAALTG 133  
 QY 121 TYYVDHLTFLQDWAHAGLNDLAVAVEPVFSDMEVKIITWGADTACGDIISGLPVSAR 180  
 Db 121 TYYVDHLTFLQDWAHAGLNDLAVAVEPVFSDMEVKIITWGADTACGDIISGLPVSAR 180

```

Db 134 TVYDHLTPLODMAHAGRLDAVAEPIFSDMVEVKITWGADTACGDIISGLPVSARR 193
QY 181 GREILLGPADNFEQGGWRLAPITAYVSOQTRGLICITSLTGRDKNOVEGEVQVSTAT 240
Db 194 GREILLGPADNFEQGGWRLAPITAYVSOQTRGLICITSLTGRDKNOVEGEVQVSTAT 253
QY 241 OSFLATCVNGVCMVTFHAGSKTLAAGPKPITQMTNTVDQDLVGMQAPPGARSMTPTCTG 300
Db 254 OSFLATCVNGVCMVTFHAGSKTLAAGPKPITQMTNTVDQDLVGMQAPPGARSMTPTCTG 313
QY 301 SSDLYLVTRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLICPSGHAAGIFPAAVCT 360
Db 314 SSDLYLVTRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLICPSGHAAGIFPAAVCT 373
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 374 RGVAKAVDFIPVESMETTMR 393

```

RESULT 3  
ABG32181  
ID ABG32181 standard; protein; 409 AA.

ABG32181;

05-NOV-2002 (first entry)

HCV protease NS2/3 (810-1206).

HCV: enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
hepatocellular carcinoma; anti-inflammatory; lauryldiethyldiamine oxide; LDAO;  
chaotropic agent; mutant; mutagen.

Hepatitis C virus.  
Synthetic.

Key Location/Qualifiers  
Peptide 398..409  
/note="Streptavidin tag"

MO200248375-A2.

20-JUN-2002.

13-DEC-2001; 2001MO-CA001796.

15-DEC-2000; 2000US-0256031P.

(BOEH) BOEHRINGER INGELHEIM CANADA LTD.

Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;

WPI, 2002-559511/64.

DR N-PSDB; ABK90406.

Novel polypeptide for screening inhibitors of non-structural proteases  
useful as therapeutic agents against hepatitis C virus, comprises full  
length non-structural protease, or its truncation.

Claim 42; Fig 1B; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-  
length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
residue amino acid 810 to 906, or having a minimal amino acid sequence  
from residue 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
NS2/3 protease. Also included are (1) a composition (C) comprising an  
isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
its truncation or a mutated sequence, where the protease is in a solution  
comprising a sufficient concentration of lauryldiethyldiamine oxide (LDAO)  
to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32181; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 (810-1206) protein, which has a C-  
CC terminal streptavidin tag  
XX

Sequence 409 AA;

Query Match 100.0%; Score 1987; DB 5; Length 409;  
Best Local Similarity 100.0%; Pred. No. 1,6e-184;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ALLTSPYKYLARLWLOYLITRYBAHQWITPPLNRRGSDAIIILTCANPELIF 60
Db 18 ALLTSPYKYLARLWLOYLITRYBAHQWITPPLNRRGSDAIIILTCANPELIF 77
QY 61 DITKLLAIFGLPLVWLOGITKVPYFAAGLIRACMLVRAAGHYVQMAFMLAALTG 120
Db 78 DITKLLAIFGLPLVWLOGITKVPYFAAGLIRACMLVRAAGHYVQMAFMLAALTG 137
QY 121 TVYDHLTPLODMAHAGRLDAVAEPIFSDMVEVKITWGADTACGDIISGLPVSARR 180
Db 138 TVYDHLTPLODMAHAGRLDAVAEPIFSDMVEVKITWGADTACGDIISGLPVSARR 197
QY 181 GREILLGPADNFEQGGWRLAPITAYVSOQTRGLICITSLTGRDKNOVEGEVQVSTAT 240
Db 198 GREILLGPADNFEQGGWRLAPITAYVSOQTRGLICITSLTGRDKNOVEGEVQVSTAT 257
QY 241 OSFLATCVNGVCMVTFHAGSKTLAAGPKPITQMTNTVDQDLVGMQAPPGARSMTPTCTG 300
Db 258 OSFLATCVNGVCMVTFHAGSKTLAAGPKPITQMTNTVDQDLVGMQAPPGARSMTPTCTG 317
QY 301 SSDLYLVTRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLICPSGHAAGIFPAAVCT 360
Db 318 SSDLYLVTRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLICPSGHAAGIFPAAVCT 377
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 378 RGVAKAVDFIPVESMETTMR 397

```

RESULT 4  
AAR82694  
ID AAR82694 standard; protein; 3010 AA.

AAR82694;

16-OCT-2003 (revised)  
DT 14-NOV-1996 (first entry)

Partial HCV non-structural polyprotein.

protease; hepatitis C virus; screening; inhibitor; proteolytic;  
identification; cleavage.

XX

OS Hepatitis C virus; Virus.  
 XX Key Location/Qualifiers  
 FH 898..1233  
 FT Protein /note="Partial proteinase; see AAR82692"  
 FT Protein 992..1907  
 FT /note="partial proteinase; see AAR82693"  
 XX JF07184648-A.  
 XX  
 XX 25-JUL-1995.  
 XX  
 XX 05-FEB-1993; 93JP-00018854.  
 XX  
 XX 07-FEB-1992; 92JP-00022657.  
 PR 18-SEP-1992; 92JP-00249240.  
 PR 04-DEC-1992; 92JP-00325303.  
 XX  
 XX (KAEN/) KAENNO K.  
 PA (SUMO) SUMITOMO METAL IND LTD.  
 PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  
 XX  
 XX WPI; 1995-287962/38.  
 DR N-PSDB; AAT03960.  
 XX  
 XX An HCV proteinase active substance - which has activity as an anti-HCV  
 PT agent and can be used to screen for proteinase inhibitors.  
 XX  
 XX Disclosure; Page 39-48; 52pp; Japanese.  
 XX  
 XX The present sequence is a partial Hepatitis C virus (HCV) polyprotein  
 CC from the non-structural region. Partial proteinase sequences (AAR82692-  
 CC 93) are contained within this sequence. The proteinases can be used as  
 CC anti-HCV agents. They can also be used to screen cpgs. for their ability  
 CC to inhibit their proteolytic activity. In this way proteinase inhibitors  
 CC can be identified. (Updated on 16-OCT-2003 to standardise OS field)  
 XX  
 XX Sequence 3010 AA;  
 SQ  
 Query Match 95.7%; Score 1902; DB 2; Length 3010;  
 Best Local Similarity 94.2%; Pred. No. 5.3e-175;  
 Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0;  
 QY 2 LTLSFYKYLARLIMLYITRVEAHLQVWIPPLNVRGGRDAIILITCAVHPELIFD 61  
 DB 828 LTLSFYKYLARLIMLYITRVEAHLQVWIPPLNVRGGRDAIILITCAVHPELIFD 887  
 QY 62 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALTGT 121  
 DB 888 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALTGT 947  
 QY 122 YVYDHLTPLODMAHAGLRDLAAVAEPVIFSDMEVKIITMGADTAACGDIISGLPVSARG 181  
 DB 948 YVYDHLTPLODMAHAGLRDLAAVAEPVIFSDMETKLTITMGADTAACGDIISGLPVSARG 1007  
 QY 182 REILIGPADNPEGQGRLLAPITAYSOQTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 241  
 DB 1008 REILIGPADNPEGQGRLLAPITAYSOQTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 1067  
 QY 242 SFLATCVMGVCMTVPHGAGSKTLAGKGPITOMYTNVDDLVGMQAPRGASMTPTCTGS 301  
 DB 1068 SFLATCVMGVCMTVPHGAGSKTLAGKGPITOMYTNVDDLVGMQAPRGASMTPTCTGS 1127  
 QY 302 SDLYVTRHADYIPVRRGDSRGLSPRVSYLKGSGGRLPCSGHAGVIFRAAVCTR 361  
 DB 1128 SDLYVTRHADYIPVRRGDSRGLSPRVSYLKGSGGRLPCSGHAGVIFRAAVCTR 1187  
 QY 362 GVAKAVDFIPVESMETTMR 380  
 DB 1188 GVAKAVDFIPVESMETTMR 1206  
 RESULT 5

AAR86822  
 ID AAR86822 standard; protein; 3010 AA.  
 XX  
 XX AAR86822;  
 AC  
 XX  
 XX 16-OCT-2003 (revised)  
 DT 16-OCT-1995 (first entry)  
 DT  
 XX HCV protein cleavable with new serine proteinase.  
 DE  
 XX  
 XX proteinase; serine; cleavage; hepatitis C virus; HCV.  
 XX  
 XX Hepatitis C virus; Virus.  
 XX  
 XX  
 XX Key Location/Qualifiers  
 FH Cleavage-site 2419..2420  
 FT /note="Serine protease cleavage site"  
 FT  
 XX  
 XX JF06315377-A.  
 XX  
 XX 15-NOV-1994.  
 XX  
 XX 06-MAY-1993; 93JP-00105666.  
 XX  
 XX 06-MAY-1993; 93JP-00105666.  
 XX  
 XX (KAEN/) KAENNO K.  
 PA (SUMO) SUMITOMO METAL IND LTD.  
 PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  
 XX  
 XX WPI; 1995-032330/05.  
 DR N-PSDB; AAG80498.  
 XX  
 XX New HCV-originated proteinase active substance - used for site-specific  
 PT cleavage by an intermolecular reaction and the purification thereof.  
 XX  
 XX Disclosure; Page 10-19; 23pp; Japanese.  
 XX  
 XX This protein from HCV (hepatitis C virus) (encoded by AAO80498) is  
 CC cleaved between amino acids 2419 and 2420, by a new serine protease.  
 CC contg. the sequence of AAR86821. The proteinase is purified as a fused  
 CC product with the dihydrofolate reductase protein by using a methotrexate  
 CC column. It can be used for the development of an inhibitor for HCV  
 CC proteinase. (Updated on 16-OCT-2003 to standardise OS field)  
 XX  
 XX Sequence 3010 AA;  
 SQ  
 Query Match 95.5%; Score 1897; DB 2; Length 3010;  
 Best Local Similarity 93.9%; Pred. No. 1.6e-174;  
 Matches 356; Conservative 12; Mismatches 11; Indels 0; Gaps 0;  
 QY 2 LTLSFYKYLARLIMLYITRVEAHLQVWIPPLNVRGGRDAIILITCAVHPELIFD 61  
 DB 828 LTLSFYKYLARLIMLYITRVEAHLQVWIPPLNVRGGRDAIILITCAVHPELIFD 887  
 QY 62 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALTGT 121  
 DB 888 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALTGT 947  
 QY 122 YVYDHLTPLODMAHAGLRDLAAVAEPVIFSDMEVKIITMGADTAACGDIISGLPVSARG 181  
 DB 948 YVYDHLTPLODMAHAGLRDLAAVAEPVIFSDMETKLTITMGADTAACGDIISGLPVSARG 1007  
 QY 182 REILIGPADNPEGQGRLLAPITAYSOQTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 241  
 DB 1008 REILIGPADNPEGQGRLLAPITAYSOQTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 1067  
 QY 242 SFLATCVMGVCMTVPHGAGSKTLAGKGPITOMYTNVDDLVGMQAPRGASMTPTCTGS 301  
 DB 1068 SFLATCVMGVCMTVPHGAGSKTLAGKGPITOMYTNVDDLVGMQAPRGASMTPTCTGS 1127  
 QY 302 SDLYVTRHADYIPVRRGDSRGLSPRVSYLKGSGGRLPCSGHAGVIFRAAVCTR 361

Db 1128 SDLYLVTRHADVPVRRRGRSGSLSPRPISTYLKSSGGPILCPGSHVVGIFRAAVCTR 1187

QY 362 GVAKAVDPIPVESMETTR 380

Db 1188 GVAKAVDPIPVESMETTR 1206

RESULT 6

AA068864

ID AA068864 standard; protein, 3010 AA.

XX AA068864;

AC 06-DEC-1995 (first entry)

DT 06-DEC-1995 (first entry)

XX Hepatitis C virus RNA helicase.

DE Hepatitis C virus RNA helicase.

XX Hepatitis C virus; HCV, non-A non-B; helicase gene; RNA helicase;

KM Baculovirus; recombinant production.

XX Hepatitis C virus.

OS Hepatitis C virus.

XX Key

Location/Qualifiers

196..198

FT Region /label= N-linked glycosylation site

FT 209..211

FT Region /label= N-linked glycosylation site

FT 234..236

FT Region /label= N-linked glycosylation site

FT 250..252

FT Region /label= N-linked glycosylation site

FT 305..307

FT Region /label= N-linked glycosylation site

FT 325..327

FT Region /label= N-linked glycosylation site

FT 417..419

FT Region /label= N-linked glycosylation site

FT 423..425

FT Region /label= N-linked glycosylation site

FT 430..432

FT Region /label= N-linked glycosylation site

FT 448..450

FT Region /label= N-linked glycosylation site

FT 532..534

FT Region /label= N-linked glycosylation site

FT 556..558

FT Region /label= N-linked glycosylation site

FT 576..578

FT Region /label= N-linked glycosylation site

FT 623..625

FT Region /label= N-linked glycosylation site

FT 645..647

FT Region /label= N-linked glycosylation site

FT 1213..1215

FT Region /label= N-linked glycosylation site

FT 1255..1257

FT Region /label= N-linked glycosylation site

FT 2041..2043

FT Region /label= N-linked glycosylation site

FT 2077..2079

FT Region /label= N-linked glycosylation site

FT 2240..2242

FT Region /label= N-linked glycosylation site

FT 2788..2790

FT Region /label= N-linked glycosylation site

XX JP06319583-A.

XX 22-NOV-1994.

XX 18-SEP-1992; 92JP-00249241.

XX 18-SEP-1992; 92JP-00249241.

PR

XX (SOYA-) SOYAKU GIUTTSU KENKUTUSHO KK.

PA WPI, 1995-040330/06.

DR N-PSDB; AA081559.

XX of hepatitis C virus helicase gene in baculovirus - useful for large

PT scale prodn. of RNA helicase.

XX Claim 1, Fig 1-4, 9pp; Japanese.

PS AA081559 encodes AA068864 hepatitis C virus (HCV) RNA helicase. The DNA

CC was used in the construction of an expression vector, which was used to

CC transform a baculovirus host. The transformed baculovirus could then be

CC used for the recombinant prodn. of HCV RNA helicase

XX Sequence 3010 AA;

SQ

Query Match 95.4%; Score 1896; DB 2; Length 3010;

Best Local Similarity 93.9%; Pred. No. 2e-174;

Matches 356; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 2 LITSPYKYLARLIMWLOVITRVEAHLOVWIPINVGSGDAILITCAVHPELIP 61

Db LITSPYKYLARLIMWLOVITRVEAHLOVWIPINVGSGDAILITCAVHPELIP 887

QY 62 ITKLLALFGEPLMVLQAGITKVPYFVRAQGLIRACMLVRRKAGSHYVQNAFMELALTG 121

Db ITKLLALFGEPLMVLQAGITKVPYFVRAQGLIRACMLVRRKAGSHYVQNAFMELALTG 947

QY 122 VYVDHLPLDQMAAG-RDLAVAVEPIFSDMEVKIITWGADPAAGDIISGLPVABRG 181

Db VYVDHLPLDQMAAG-RDLAVAVEPIFSDMEVKIITWGADPAAGDIISGLPVABRG 1007

QY 948 VYVDHLPLDQMAAG-RDLAVAVEPIFSDMEVKIITWGADPAAGDIISGLPVABRG 1007

Db 182 REILGPADNPEGGMELAPITAYSGQTRGLGCIITSLTGRDKQVGEVGVSTAQ 241

Db KEILGPADNPEGGMELAPITAYSGQTRGLGCIITSLTGRDKQVGEVGVSTAQ 1067

QY 242 SFLATCNVGVCMVYFHAGSKTLGAPKPTOMVTVNDOLVGMQAPRARSMTPTCTGS 301

Db SFLATCNVGVCMVYFHAGSKTLGAPKPTOMVTVNDOLVGMQAPRARSMTPTCTGS 1127

QY 1068 SFLATCNVGVCMVYFHAGSKTLGAPKPTOMVTVNDOLVGMQAPRARSMTPTCTGS 1127

Db SDLYLVTRHADVPVRRRGRSGSLSPRPISTYLKSSGGPILCPGSHVVGIFRAAVCTR 1187

QY 362 GVAKAVDPIPVESMETTR 380

Db 1188 GVAKAVDPIPVESMETTR 1206

RESULT 7

AA040223

ID AA040223 standard; protein, 768 AA.

XX AA040223;

AC 21-FEB-1994 (first entry)

DT 21-FEB-1994 (first entry)

XX Recombinant hepatitis C virus genomic protein.

DE Recombinant hepatitis C virus genomic protein.

XX Hepatitis; HCV; virus; screening; antiviral drugs.

OS Hepatitis C virus.

XX Key

Location/Qualifiers

10

FT Misc-difference /note= "UUA encodes Ile."

FT Misc-difference 81..82

FT /note= "Nucleotide sequence encodes another Gly"

FT Duplication 528..768

FT /note= "Duplication of 241 amino acids at start of protein sequence."

FT Misc-difference 537 /note= "UUA encodes Ile."

XX JP05192160-A.

XX 03-AUG-1993.

PD 20-JAN-1992; 92JP-00028833.

XX 20-JAN-1992; 92JP-00028833.

PR 20-JAN-1992; 92JP-00028833.

XX (BANY ) BANYU PHARM CO LTD.

PA WPI; 1993-277474/35.

DR N-PSDB; AAQ48215.

XX

PT Hepatitis C virus genomic RNA, cDNA and polypeptide - used for screening

PT hepatitis C virus-specific antiviral drugs.

XX

PS Claim 9; Page 4-6; 14pp; Japanese.

XX

CC The protein is useful in the screening of HCV-specific antiviral drugs.

CC HCV cDNA was cloned from plasma. Plasmids pSR3241 and pSR2541 were

CC prepared using the cDNA and plasmid pSR3241 was used to transform a COS-1

CC cell

SO Sequence 768 AA;

Query Match 95.0%; Score 1888; DB 2; Length 768;

Best Local Similarity 93.2%; Pred. No. 1.8e-174;

Matches 354; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

QY 1 ALTLSPYKVLARLIMWLYITVEAHLOVWIPINRGGRDAIILTCVAHPELIF 60

DB 88 ALTLSPYKVLARLIMWLYITVEAHLOVWIPINRGGRDAIILTCVAHPELIF 147

QY 61 DITKLALIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 120

DB 148 DITKLALIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 207

QY 121 TYVYDHLTPLODMANAGRLDVAVAEPIVSDMEVKIITWGADTAACGDIISGLPVSARR 180

DB 208 TYVYDHLTPLODMANAGRLDVAVAEPIVSDMEVKIITWGADTAACGDIISGLPVSARR 267

QY 181 GREIILGPADNPEGQWRLLAPITAYSQOTRGILGCIITSLGRDNQVGEVQVAVTAT 240

DB 268 GREIILGPADNPEGQWRLLAPITAYSQOTRGILGCIITSLGRDNQVGEVQVAVTAT 327

QY 241 QSFPLATCVNGVCMVTFHGAAGSKTLGPKGPIITQYTNVDDVLGMAQAPGARSMTPCTCG 300

DB 328 QSFPLATCVNGVCMVTFHGAAGSKTLGPKGPIITQYTNVDDVLGMAQAPGARSMTPCTCG 387

QY 301 SSDLIVLTHADVITPRRRGDSRGSILSPRPVSYLKSGSGGPLCPSGHAVGIFRAAVCT 360

DB 388 SSDLIVLTHADVITPRRRGDSRGSILSPRPVSYLKSGSGGPLCPSGHAVGIFRAAVCT 447

QY 361 RGVAKAVDEIPVESMETTKR 380

DB 448 RGVAKAVDEIPVESMETTKR 467

RESULT 8

ABG30601

ID ABG30601 standard; protein; 2201 AA.

XX

AC ABG30601;

XX

DT 21-OCT-2002 (first entry)

XX

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.

XX

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

XX

KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.

XX Hepatitis C virus.

OS Synthetic.

XX

XX Key Location/Qualifiers

FT Misc-difference 882

FT /label= Arg, Lys

FT Misc-difference 2183

FT /note= "Wild type Met substituted by Thr"

XX

FN WO200252015-A2.

XX

PD 04-JUL-2002.

XX

PF 20-DEC-2001; 2001WO-CR001843.

XX

PR 22-DEC-2000; 2000US-0257857P.

XX

PA (BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.

XX

PI Kukolj G, Pause A;

XX

DR WPI; 2002-575382/61.

XX

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which

PT possess enhanced transduction or replication efficiency, useful for

PT evaluating potential inhibitors of HCV replication.

XX

PS Claim 3; Page; 140pp; English.

XX

CC The invention describes a self-replicating hepatitis C virus (HCV)

CC polynucleotide molecule comprising a 5'-non translated region (NTR),

CC where guanine at position 1 is substituted for adenine, a HCV polypeptin

CC region coding for a HCV polypeptin; and a 3'-NTR region. The self-

CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating

CC potential inhibitors of HCV replication. The HCV RNA molecule is also

CC useful for efficiently establishing cell culture replication. The self-

CC replicating polynucleotide molecule contains a 5'-NTR, where G at

CC position 1 is substituted for A, and therefore provides an alternative to

CC existing systems comprising a self-replicating HCV RNA molecule that, in

CC conjunction with mutations in the HCV non-structural region, such as the

CC G12042/C/R mutations, transduces and/or replicates with greater

CC efficiency. This amino acid sequence represents a mutant of the hepatitis

CC C virus replicon ABG312 and contains the viral protease NS2/3, protease

CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:

CC This sequence does not appear in the specification but has been created

CC from the wild type sequence shown in ABG30580 using information given in

CC the claims of the invention

XX

SO Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;

Best Local Similarity 93.4%; Pred. No. 9.9e-174;

Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMWLYITVEAHLOVWIPINRGGRDAIILTCVAHPELIFD 61

DB 19 LTLSPYKVLARLIMWLYITVEAHLOVWIPINRGGRDAIILTCVAHPELIFD 78

QY 62 IYKLLAIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 121

DB 79 IYKLLAIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 138

QY 122 TYVYDHLTPLODMANAGRLDVAVAEPIVSDMEVKIITWGADTAACGDIISGLPVSARRG 181

DB 139 TYVYDHLTPLODMANAGRLDVAVAEPIVSDMEVKIITWGADTAACGDIISGLPVSARRG 198

QY 182 REIILGPADNPEGQWRLLAPITAYSQOTRGILGCIITSLGRDNQVGEVQVAVTATQ 241

DB 199 REIILGPADNPEGQWRLLAPITAYSQOTRGILGCIITSLGRDNQVGEVQVAVTATQ 258

QY 242 SFLATCVNGVCMVTFHGAAGSKTLGPKGPIITQYTNVDDVLGMAQAPGARSMTPCTCGS 301



DB 259 SFLATCVNGVCMVTYHAGSKTLAGPKPITOMTYNDODLVGMQAPPGARSILPCTCGS 318  
 QY 302 SDLYLVTRHADVIPIVRRRGRSGSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTR 361  
 DB 319 SDLYLVTRHADVIPIVRRRGRSGSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTR 378  
 QY 362 GYAKAVDFIPVESMETTMR 380  
 DB 379 GYAKAVDFIPVESMETTMR 397  
 RESULT 9  
 ABG30591  
 ID ABG30591 standard; protein; 2201 AA.  
 AC ABG30591;  
 XX 21-OCT-2002 (first entry)  
 DT Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 DE  
 XX Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 751  
 FT /note= "Wild type Ser substituted by Gly"  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 FT  
 FN WO200252015-A2.  
 XX 04-JUL-2002.  
 PD 20-DEC-2001; 2001WO-CA001843.  
 PF 22-DEC-2000; 2000US-0257857P.  
 PR (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PA Kukulj G, Pause A;  
 PI WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apk12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC this sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 XX Sequence 2201 AA:

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
 Best Local Similarity 93.4%; Pred. No. 9,9e-174;  
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;  
 QY 2 LTLSPYKXVLLARLIMWLOYLITRVEAHQWVWIPPLNVRGGRDAIILLTCVHPELIFD 61  
 DB 19 LTLSPYKXVLLARLIMWLOYLITRVEAHQWVWIPPLNVRGGRDAIILLTCVHPELIFD 78  
 QY 62 ITKLILAIFGIMWLOIGITKVPFVRAOGLIRACMLVRAAGHYVQMAFMKLAALTGT 121  
 DB 79 ITKLILAIFGIMWLOIGITKVPFVRAOGLIRACMLVRAAGHYVQMAFMKLAALTGT 138  
 QY 122 YVYDHLTPLODMAHAGRDIAVAVEPIFSDMEVKIITWGADTAACGDIISGLFVSARRG 181  
 DB 139 YVYDHLTPLODMAHAGRDIAVAVEPIFSDMEVKIITWGADTAACGDIISGLFVSARRG 198  
 QY 182 REILGPADNFEQGMRLAPITAYSOOTRGLACITTSITGRDKNVBEQVAVSTATQ 241  
 DB 199 REILGPADNFEQGMRLAPITAYSOOTRGLACITTSITGRDKNVBEQVAVSTATQ 258  
 QY 242 SFLATCVNGVCMVTYHAGSKTLAGPKPITOMTYNDODLVGMQAPPGARSMT PCTCGS 301  
 DB 259 SFLATCVNGVCMVTYHAGSKTLAGPKPITOMTYNDODLVGMQAPPGARSILPCTCGS 318  
 QY 302 SDLYLVTRHADVIPIVRRRGRSGSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTR 361  
 DB 319 SDLYLVTRHADVIPIVRRRGRSGSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTR 378  
 QY 362 GYAKAVDFIPVESMETTMR 380  
 DB 379 GYAKAVDFIPVESMETTMR 397  
 RESULT 10  
 ABG30600  
 ID ABG30600 standard; protein; 2201 AA.  
 AC ABG30600;  
 XX 21-OCT-2002 (first entry)  
 DT Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.  
 DE  
 XX Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 FT Misc-difference 1357  
 FT /note= "Wild type Pro substituted by Leu"  
 FT  
 FN WO200252015-A2.  
 XX 04-JUL-2002.  
 PD 20-DEC-2001; 2001WO-CA001843.  
 PF 22-DEC-2000; 2000US-0257857P.  
 PR (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PA Kukulj G, Pause A;  
 PI WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.

XX Claim 3, Page: 140pp, English.

XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polypeptide  
XX region coding for a HCV polypeptide and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in  
XX conjunction with mutations in the HCV non-structural region, such as the  
XX G(2042)/C/R mutations, transduces and/or replicates with greater  
XX efficiency. This amino acid sequence represents a mutant of the hepatitis  
XX C virus replicon ABG30581 and contains the viral protease NS2/3, protease  
XX complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
XX this sequence does not appear in the specification but has been created  
XX from the wild type sequence shown in ABG30580 using information given in  
XX the claims of the invention

XX Sequence 2201 AA:

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLSPYKYLRLIWMLOYLITRVEAHQWIPPLNVRGGRDAIILITCAVPELIFD 61  
DB 19 LTLSPHYKLFARLIMWLOYFITRAEHLQWIPPLNVRGGRDAIILITCAIHEPILFT 78  
QY 62 ITKLALIFGLPLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 121  
DB 79 ITKLALILGLPLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 138  
QY 122 YVVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 181  
DB 139 YVVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 198  
QY 182 REILGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVNSTAQ 241  
DB 199 REILGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVNSTAQ 258  
QY 242 SFPLATCVNGVCMVFHAGSKTLGAPKPIITQMTNVDDLVGMQAPPARASLTPTCTGS 301  
DB 259 SFPLATCVNGVCMVFHAGSKTLGAPKPIITQMTNVDDLVGMQAPPARASLTPTCTGS 318  
QY 302 SDLYLTVTRADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361  
DB 319 SDLYLTVTRADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 378  
QY 362 GVAKAVDFIPVESMETTMR 380  
DB 379 GVAKAVDFIPVESMETTMR 397

RESULT 11

ABG30581  
ID ABG30581 standard; protein; 2201 AA.

XX ABG30581;

XX 21-OCT-2002 (first entry)

XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.

XX Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor;

XX cell culture replication; NS2/3; NS3/4; NS3; NS5B.

XX Hepatitis C virus.

XX WO200252015-A2.

XX 04-JUL-2002.

XX 20-DEC-2001; 2001WO-CA001843.

XX 22-DEC-2000; 2000US-0257857P.

XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.

XX Kukulj G, Pause A;

XX WPI: 2002-575382/61.

XX N-PSDB: ABK88573.

XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
XX possess enhanced transduction or replication efficiency, useful for  
XX evaluating potential inhibitors of HCV replication.

XX Disclosure: Page 49-58; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polypeptide  
XX region coding for a HCV polypeptide and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in  
XX conjunction with mutations in the HCV non-structural region, such as the  
XX G(2042)/C/R mutations, transduces and/or replicates with greater  
XX efficiency. This amino acid sequence is encoded by the hepatitis C virus  
XX replicon ABG30581 and contains the viral protease NS2/3, protease complex  
XX NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

XX Sequence 2201 AA:

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLSPYKYLRLIWMLOYLITRVEAHQWIPPLNVRGGRDAIILITCAVPELIFD 61  
DB 19 LTLSPHYKLFARLIMWLOYFITRAEHLQWIPPLNVRGGRDAIILITCAIHEPILFT 78  
QY 62 ITKLALIFGLPLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 121  
DB 79 ITKLALILGLPLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 138  
QY 122 YVVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 181  
DB 139 YVVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 198  
QY 182 REILGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVNSTAQ 241  
DB 199 REILGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVNSTAQ 258  
QY 242 SFPLATCVNGVCMVFHAGSKTLGAPKPIITQMTNVDDLVGMQAPPARASLTPTCTGS 301  
DB 259 SFPLATCVNGVCMVFHAGSKTLGAPKPIITQMTNVDDLVGMQAPPARASLTPTCTGS 318  
QY 302 SDLYLTVTRADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361  
DB 319 SDLYLTVTRADVIPIVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 378  
QY 362 GVAKAVDFIPVESMETTMR 380  
DB 379 GVAKAVDFIPVESMETTMR 397

RESULT 12

ABG30593

ID ABG30593 standard; protein; 2201 AA.  
XX  
AC ABG30593;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #4.  
XX  
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 882  
FT /label= Arg, Lys  
FT Misc-difference 892  
FT /note= "Wild type Leu substituted by Phe"  
XX  
XX MO200252015-A2.  
XX  
PD 04-JUL-2002.  
XX  
PF 20-DEC-2001; 2001MO-CA001843.  
XX  
PR 22-DEC-2000; 2000US-0257857P.  
XX  
PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
PI Kukolj G, Pause A;  
XX  
XX WPI; 2002-575382/61.  
XX  
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.  
XX  
XX Claim 3; Page; 140pp; English.  
XX  
PS The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polyprotein  
XX region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in  
XX conjunction with mutations in the HCV non-structural region, such as the  
XX G(2042)/C/R mutations, transduces and/or replicates with greater  
XX efficiency. This amino acid sequence represents a mutant of the hepatitis  
XX C virus replicon ApeK12 and contains the viral protease NS2/3, protease  
XX complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
XX This sequence does not appear in the specification but has been created  
XX from the wild type sequence shown in ABG30580 using information given in  
XX the claims of the invention  
XX  
SQ Sequence 2201 AA;  
XX  
Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;  
XX  
QY 2 LITTSPPYKULALITMLQVITRVEAHLOVWIPPLNVRGGRPAIITLTCVAPHEIIPD 61  
DB 19 LITTSPPYKULALITMLQVITRVEAHLOVWIPPLNVRGGRPAIITLTCVAPHEIIPD 61  
QY 62 ITKILALFSPPLNVQAGITKVPFVRAQGHIRACMLVRKKAAGHYVQMAFMKLAALTGT 121  
DB 79 ITKILALFSPPLNVQAGITKVPFVRAQGHIRACMLVRKKAAGHYVQMAFMKLAALTGT 138

QY 122 YVYDHLTPLOPMARAGLRDLAVAVEPIFSDMEKXITMGADTAAGDITISGLPVASARG 181  
DB 139 YVYDHLTPLOPMARAGLRDLAVAVEPIFSDMEKXITMGADTAAGDITISGLPVASARG 198  
QY 182 REILGSPADNPEGQGWRLAPITAYSOQTRGLLCIITSLTGRDKNVESEVQVASTATQ 241  
DB 199 REILGSPADNPEGQGWRLAPITAYSOQTRGLLCIITSLTGRDKNVESEVQVASTATQ 258  
QY 242 SFLATCVNGVCMVTFHGAQSKTLGAPKPIITQMTTNDQDLYGQAPPGARSMTPTCGS 301  
DB 259 SFLATCVNGVCMVTFHGAQSKTLGAPKPIITQMTTNDQDLYGQAPPGARSMTPTCGS 318  
QY 302 SDLYVTFRHADVTPRRRGDSRGLSPRPVSYLKSGSGPLCPGSHAVGIFPAAVCTR 361  
DB 319 SDLYVTFRHADVTPRRRGDSRGLSPRPVSYLKSGSGPLCPGSHAVGIFPAAVCTR 378  
QY 362 GVAKAVDEPIPVESMETTWR 380  
DB 379 GVAKAVDEPIPVESMETTWR 397  
XX  
XX RESULT 13  
XX ABG30582  
ID ABG30582 standard; protein; 2201 AA.  
XX  
XX ABG30582;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.  
XX  
XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.  
XX  
OS Hepatitis C virus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 882  
FT /note= "Wild type Lys substituted by Lys or Arg"  
FT Misc-difference 1233  
FT /note= "Wild type Gly substituted by Cys"  
XX  
XX MO200252015-A2.  
XX  
PD 04-JUL-2002.  
XX  
PF 20-DEC-2001; 2001MO-CA001843.  
XX  
PR 22-DEC-2000; 2000US-0257857P.  
XX  
PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
PI Kukolj G, Pause A;  
XX  
XX WPI; 2002-575382/61.  
XX N-PSDB; ABK88574.  
XX  
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.  
XX  
XX Disclosure; Page 59-69; 140pp; English.  
XX  
XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polyprotein  
XX region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in

conjunction with mutations in the HCV non-structural region, such as the G2042/C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence is encoded by the hepatitis C virus replicon Apgk12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this sequence has been created from replicon Apgk12 shown in ABG30581

Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

2 LLLTSPYKVLARLIMWLQYLITRVAHLQWIPUNVSGRDAIILLTCVAHPLEIFD 61  
19 LLLTSPHYKFLFARLIMWLQYFIRABAHQWIPUNVSGRDAVILLTCALHPLEIF 78  
62 ITKLILAFGLMVLQAGITKVPFVFAOGLIPACMLVRRAGGHYQVAFMLAALTGT 121  
79 ITKLILAFGLMVLQAGITKVPFVFAOGLIPACMLVRRAGGHYQVAFMLAALTGT 138  
122 YVVDHLTPLODMAHAGRLAVALVEPVIFSDMEVKIITWGAADTAACGDIILGLPVSAARG 181  
139 YVVDHLTPLODMAHAGRLAVALVEPVIFSDMEVKIITWGAADTAACGDIILGLPVSAARG 198  
182 REILGPAENFEGGWMFLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 241  
199 REILGPAENFEGGWMFLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 258  
242 SFATCVNGVCMVTFHAGSKTLAGPKGPTTOMYTNVDOLVGMQAPPGARSLTPTCGS 301  
259 SFATCVNGVCMVTFHAGSKTLAGPKGPTTOMYTNVDOLVGMQAPPGARSLTPTCGS 318  
302 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361  
319 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 378  
362 GVAKAVDFIVESMETTMR 380  
379 GVAKAVDFIVESMETTMR 397

RESULT 14

ABG30580 ID ABG30580 standard; protein; 2201 AA.

AC ABG30580;  
DT 21-OCT-2002 (first entry)  
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #3.  
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
KW cell culture replication; NS2/3, NS3/4, NS3; NS5B.  
OS Hepatitis C virus.  
XX  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 882 /note= "Encoded by ARG"  
XX  
XX WO200252015-A2.  
XX PD 04-JUL-2002.  
XX PF 20-DEC-2001; 2001WO-CA001843.  
XX PR 22-DEC-2000; 2000US-0257857P.  
XX  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX PA  
XX PI Kuko1j G, Pause A;  
XX

DR WPI; 2002-575382/61.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.

PS Disclosure; Page 69-74; 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)  
CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
CC useful for efficiently establishing cell culture replication. The self-  
CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
CC position 1 is substituted for A, and therefore provides an alternative to  
CC existing systems comprising a self-replicating HCV RNA molecule that, in  
CC conjunction with mutations in the HCV non-structural region, such as the  
CC G2042/C/R mutations, transduces and/or replicates with greater  
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
CC replicon Apgk12 and contains the viral protease NS2/3, protease complex  
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

2 LLLTSPYKVLARLIMWLQYLITRVAHLQWIPUNVSGRDAIILLTCVAHPLEIFD 61  
19 LLLTSPHYKFLFARLIMWLQYFIRABAHQWIPUNVSGRDAVILLTCALHPLEIF 78  
62 ITKLILAFGLMVLQAGITKVPFVFAOGLIPACMLVRRAGGHYQVAFMLAALTGT 121  
79 ITKLILAFGLMVLQAGITKVPFVFAOGLIPACMLVRRAGGHYQVAFMLAALTGT 138  
122 YVVDHLTPLODMAHAGRLAVALVEPVIFSDMEVKIITWGAADTAACGDIILGLPVSAARG 181  
139 YVVDHLTPLODMAHAGRLAVALVEPVIFSDMEVKIITWGAADTAACGDIILGLPVSAARG 198  
182 REILGPAENFEGGWMFLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 241  
199 REILGPAENFEGGWMFLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 258  
242 SFATCVNGVCMVTFHAGSKTLAGPKGPTTOMYTNVDOLVGMQAPPGARSLTPTCGS 301  
259 SFATCVNGVCMVTFHAGSKTLAGPKGPTTOMYTNVDOLVGMQAPPGARSLTPTCGS 318  
302 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361  
319 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 378  
362 GVAKAVDFIVESMETTMR 380  
379 GVAKAVDFIVESMETTMR 397

RESULT 15

ABG30587 ID ABG30587 standard; protein; 2201 AA.

AC ABG30587;  
DT 21-OCT-2002 (first entry)  
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #7.  
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
KW cell culture replication; NS2/3, NS3/4, NS3; NS5B.  
OS Hepatitis C virus.